University of Alberta

Mitochondrial DNA Polymorphisms as Genetic Markers for Endometrial Carcinoma and Multiple Myeloma

by

Karmon Eve Helmle

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Medical Sciences – Medical Genetics

Edmonton, Alberta

Spring 2004

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-96484-1
Our file Notre référence
ISBN: 0-612-96484-1

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract:

The presence of homoplasmic tumour-specific variants within the mitochondrial DNA and the expansion of a C-tract at position 303 have been proposed as genetic cancer markers. A subcloning-based sequencing approach was employed to gain insight into the presence of these markers in multiple myeloma and endometrial cancer, as well as in the normal population. In contrast to the findings in the current literature, this assay allowed me to demonstrate that both cancerous and normal individuals display a high level of polymorphism, C-tract length variance, and heteroplasmy, suggesting these markers may not be appropriate for cancer diagnostics. I also developed a rapid and sensitive means of DNA mutation analysis through integration of several processes on a microfluidic chip-based platform, which represents a significant advance in the field of microfluidic-based molecular diagnostics.

Acknowledgements

I would like to thank Dr. D. Moira Glerum and Dr. Christopher Backhouse for their guidance and support, as well as the members of my advisory committee, Dr. Susan Andrew and Dr. Linda Pilarski for their thoughts and advice. I would also like to thank Dr. Susan Andrew, Dr. Linda Pilarski, Dr. Diane Cox and Dr. Paul Goodfellow for supplying the DNA samples used in this study. I am grateful to Patricia Taylor for her wealth of ideas and assistance with the microfluidic chip assays. I would like to thank the Natural Sciences and Engineering Research Council of Canada and the Canadian Genetic Disease Network for financial support throughout my graduate training.

I would like to extend my thanks to my family for their endless love and support through the years, for believing I could be successful at anything I set my mind to and for giving me the means of achieving that success. Finally, I would like to thank Anthony Dechant for his technical and statistical expertise, the long nights of analyzing data and drawing figures, his infinite patience and continual encouragement.

Table of Contents

Chapter 1: Introduction **1**

Mitochondrial DNA 2

Structure 2

Replication 6

Transcription and Translation 9

DNA Mutation Rate and Repair 12

Mitochondrial Genetics 16

Maternal Inheritance 16

Homoplasmy and Heteroplasmy 17

Mitochondrial DNA Mutation and Disease 24

Cancer: Endometrial Cancer and Multiple Myeloma 26

Endometrial Cancer 26

Multiple Myeloma 27

Mitochondrial DNA and Cancer 28

Microfluidic Chip Technology 35

Summary 38

Chapter 2: Materials and Methods **40**

Mitochondrial DNA Specimens 41

Cell Culture 41

DNA Purification from Cultured Cells 44

PCR Amplification of HVI and HVII Mitochondrial DNA Fragments 45

Ligation and Transformation of Mitochondrial DNA PCR Products 47

Plasmid DNA Purification 48

Plasmid DNA Purification 49

Li-Cor Sequencing 50

Microfluidic Chip Experiments 51

Chapter 3: Results 53

Analysis of D-loop Status in Normal Individuals 54

Normal Samples are Highly Polymorphic 54

Normal Samples are Primarily Heteroplasmic 59

Analysis of D-loop Status in Cancer Samples 63

The Level of Polymorphism in Cancer Samples is Similar to Normal Controls 63

The Degree of Homoplasmy is Comparable in Cancer Samples and Normal Controls 69

The C303 Polymorphism Varies Considerably in Normal Controls 70

Development of Microfluidic Chip-Based mtDNA Assays 72

Restriction Digestion and On-Chip Labelling 72

Heteroduplex Analysis 73

Chapter 4: Discussion 81

Mitochondrial DNA Variation in Normal Samples 82

Mitochondrial DNA and Cancer 88

C303 Expansion as a Carcinogenic Marker 93

Microfluidic Chip Assays 95

Summary 96

References 98

List of Tables

Table 2-1. Endometrial carcinoma samples 42

Table 2-2. Multiple myeloma cancer samples 43

Table 3-1. Unreported polymorphisms 57

Table 3-2. Samples re-amplified with AmpliTaq Gold 60

Table 3-3. Incidence of homoplasmy in normal, endometrial and multiple myeloma samples 62

Table 3-4. HVI sequence results for matched normal and endometrial cancer samples 66

Table 3-5. HVII sequence results for matched normal and endometrial cancer samples 67

List of Figures

Figure 1-1. Schematic of the human mitochondrial genome 3

Figure 1-2. Simple cross microfluidic chip design 37

Figure 2-1. Schematic of the D-loop and PCR primers within HVI and HVII 46

Figure 2-2. Schematic of the pGEM-T Easy PCR cloning vector 48

Figure 3-1. Frequency of mtDNA polymorphism in endometrial, multiple myeloma and normal subclones 55

Figure 3-2. Schematic of sequencing results for 10 subclones compared to 50 subclones for a normal blood control 58

Figure 3-3. C303-tract length in normal, endometrial, and multiple myeloma subclones 71

Figure 3-4. On-chip restriction digest with sytox orange labeling 74

Figure 3-5. Heteroduplex analysis of a 50:50 population of mtDNA restriction digest products 77

Figure 3-6. Heteroduplex analysis of HVI subclones 6-2 and 7-1 78

Figure 3-7. Heteroduplex analysis of HVII subclones 4-9 and 5-8 80

List of Abbreviations

8-oxodG	8-oxodeoxyguanosine
A	Adenine
ATP	Adenosine Triphosphate
bp	Base pair
C	Cytosine
C10	C303 tract length of 10 consecutive cytosines
C303	C-tract length polymorphism beginning at nt 303
C7	C303 tract length of 7 consecutive cytosines
C8	C303 tract length of 8 consecutive cytosines
C9	C303 tract length of 9 consecutive cytosines
CPEO	Chronic progressive external ophthalmoplegia
del	Deletion
DNA	Deoxyribonucleic acid
G	Guanine
HVI	Hypervariable region I
HVII	Hypervariable region II
ins	Insertion
kDa	KiloDalton
LHON	Leber's hereditary optic neuropathy
MELAS	Mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes
MERRF	Myoclonic epilepsy with ragged red fibers

mRNA	Messenger RNA
MSI	Microsatellite instability
mtDNA	Mitochondrial DNA
NARP	Neurogenic weakness, ataxia, and retinitis pigmentosa
nt	Nucleotide
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
T	Thymine
tRNA	Transfer RNA

Chapter 1: Introduction

Mitochondrial DNA

Structure

Mammalian cells are comprised of two inter-dependent genetic systems – one housed in the nucleus and the other in the mitochondria, each governed by distinct structure, genetics and inheritance. While greater than 90% of mitochondrial proteins are encoded by the nucleus, a small subset of genes essential for energy production are encoded by the mitochondrial DNA (Neupert, 1997). Mitochondrial DNA (mtDNA) was first identified in the 1960's, but it was not until 1981 that the human mtDNA sequence was determined (Figure 1-1) (Anderson *et al.*, 1981). Each mtDNA is a double-stranded 16,569 base pair (bp) closed, circular molecule. Comprising up to 1% of the total DNA of a cell, the mtDNA encodes 22 transfer RNAs (tRNAs), two ribosomal RNAs (rRNAs) and 13 polypeptides involved in cellular respiration (Anderson *et al.*, 1981; Fernandez-Silva *et al.*, 2003). The genetic organization of the mtDNA is highly conserved among mammals (Wolstenholme, 1992).

The two rRNAs (16S and 12S) as well as the 22 tRNAs provide some of the necessary components for mitochondrial mRNA translation and protein synthesis, while the 13 polypeptides constitute subunits of the four oxidative phosphorylation complexes housed within the inner membrane of the mitochondrion. An electron transport chain couples the transfer of electrons to the transfer of protons from the mitochondrial matrix to the intermembrane space. Of the four complexes required for electron transport, only complex II requires no genetic contribution from the mtDNA. Complex V, the ATP synthase, utilizes the mitochondrial proton gradient to synthesize ATP (Smeitink *et al.*, 2001).

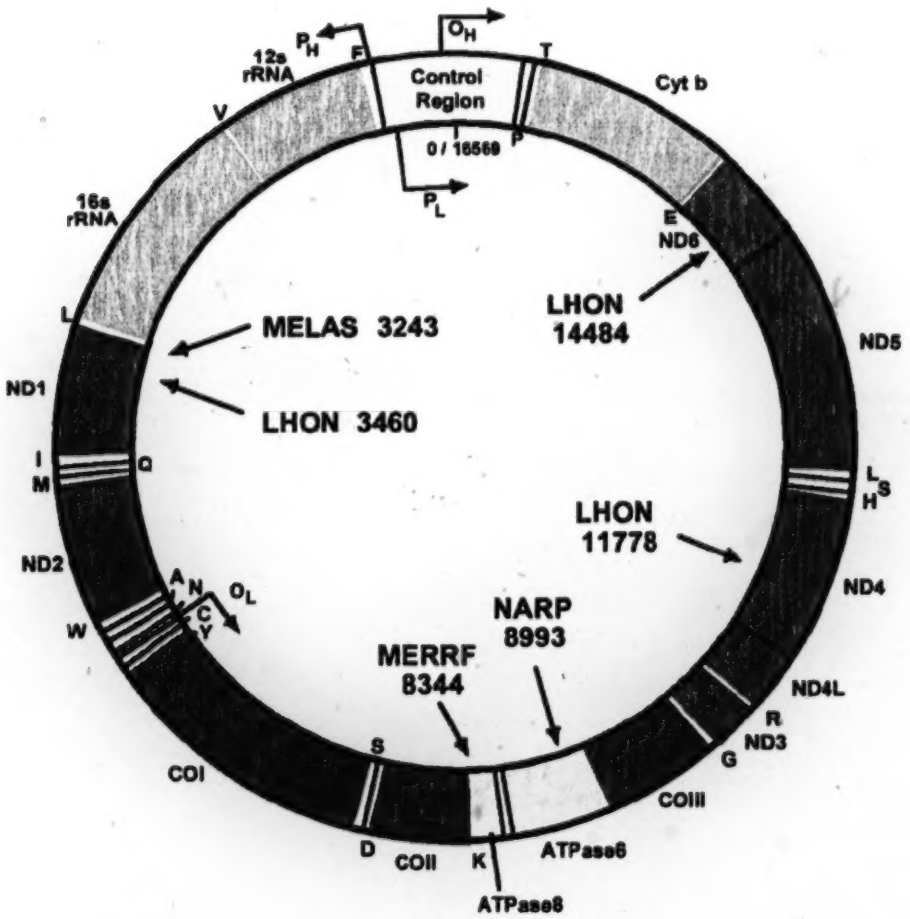

Figure 1-1. Schematic of the human mitochondrial genome. The human mtDNA is an approximately 16.5 kb circular molecule consisting of 13 protein coding genes, 22 transfer RNAs and two ribosomal RNAs. The D-loop control region encompasses two hypervariable regions, spanning nt 16024-16383 and nt 57-372. The heavy and light strand origins of replication (O_H and O_L) as well as the heavy and light strand promoter sequences (P_H and P_L) are indicated by hooked arrows. Common mutations associated with mitochondrial disease and their nucleotide positions are indicated by arrows. Source of figure: http://www.mitomap.org.

The two mtDNA strands differ in their base composition; one strand displays an abundance of guanine residues while the other strand has an abundance of complementary cytosine residues. As cytosine possesses a comparatively lower molecular weight compared to guanine, the two strands are called "heavy" and "light," reflected by the speed of their migration in a cesium chloride gradient. In addition to disparity in terms of molecular weight, the heavy and light strands are also disproportionate with regard to their genetic content, with the heavy strand serving as the template for both rRNAs and the majority of polypeptides and tRNAs, and the light strand providing the template for only one polypeptide and eight tRNAs (Wallace, 1992a).

Mitochondria likely originated from an endosymbiotic relationship between an α-purple bacterium and an ancient eukaryotic cell (Taanman, 1999). A prokaryotic origin of the mitochondria would explain its circular structure, lack of histones or other proteins involved in higher order chromatin organization, and the autonomous nature of replication, transcription and translation of mtDNA and mitochondrial proteins. Progress in areas of bioinformatics and the Human Genome Project has allowed in-depth analysis of the human nuclear genome, within which nearly 300 fragments of the mitochondrial genome, ranging in size from 100 bp to nearly 15 kb, have been inserted over the course of evolution. The pattern and organization of these "nuclear pseudogenes" has been recently characterized (Woischnik and Moraes, 2002). Surprisingly, all regions of the mtDNA genome are not equally represented in the nucleus, with mtDNA from the D-loop control region left relatively under-represented (Mourier *et al.*, 2001). The recognition of the presence of nuclear pseudogenes in the genome has required that the study of mtDNA

become more stringent in order to exclude the possibility of inadvertently investigating transferred rather than authentic mtDNA (Parfait *et al.*, 1998).

The evolutionary transfer of genetic content from the mtDNA to the nucleus has resulted in a highly organized and compact mitochondrial genome (Attardi and Schatz, 1988). The "minimalist" organization of human mtDNA is apparent in the lack of introns and intergenic regions within the molecule. The mitochondrial genes themselves are located contiguously within the DNA molecule, and the reading frames of ATPase 6 and ATPase 8 as well as ND4 and ND4L overlap, introducing an additional level of complexity to mtDNA organization (Taanman, 1999).

Only two regulatory regions exist within the mtDNA. The first and smaller non-coding region consists of approximately 30 nucleotides, located at nt 5271 - 5298 (numbered according to Anderson *et al.*, 1981) within a cluster of tRNA genes, and contains the origin of replication for the light strand (Fernandez-Silva *et al.*, 2003). The second and larger non-coding region of the mtDNA is the displacement loop (D-loop) regulatory region. The D-loop is triple stranded in nature, as a result of base-pairing between a newly created heavy strand DNA segment which base pairs with the complementary parental light strand, thereby displacing the parental heavy strand (Shadel and Clayton, 1997). The D-loop is approximately 1.1 kb, found between nt 16024 and nt 576, and contains the origin of replication for the heavy strand, as well as the two promoters required for transcription of polycistronic RNA from both the heavy and light strands (Anderson *et al.*, 1981). Also contained within the D-loop are three highly conserved regions, conserved sequence blocks I, II and III at nt 213-235, nt 299-315 and

nt 346-363, respectively, that have remained primarily unchanged throughout evolution (Wallace, 1992a).

The conserved sequence blocks are found within a larger, genetically unstable region. Two regions of the mtDNA, called hypervariable regions I and II, show the greatest variation in both base pair length and sequence composition, and are found at nt 16024-16383 and nt 57-372, respectively (Fernandez-Silva *et al.*, 2003; Shadel and Clayton, 1997; Wallace and Lott, 2003). The elevated mutation rate displayed by the hypervariable regions make them useful in studying human historical biogeography and population genetics, as well as in forensic and legal identification. Although it is widely accepted that these hypervariable regions reflect mutation hotspots, much debate exists regarding the exact rate and pattern in which the mutations occur (Meyer *et al.*, 1999). The rate of mutation within the hypervariable regions appears to be sequence and/or position dependent, with transitions occurring much more frequently than transversions (Aquadro and Greenberg, 1983; Wakeley, 1994).

Replication

Mitochondrial DNA is replicated in the mitochondrial matrix from the two unidirectional origins mentioned above, and initiation of this process is independent of the cell cycle and distinct from nuclear DNA replication (Bogenhagen and Clayton, 1977; Fernandez-Silva *et al.*, 2003; Wallace, 1992a). MtDNA replication is partially reliant on transcription, as replication of the heavy strand requires short RNA primers that are obtained through processing of transcripts originating from the promoter region of the light strand (Clayton, 1992). Replication, therefore, begins with mitochondrial RNA polymerase transcribing an immature RNA primer from the light strand promoter of the

mtDNA. The immature primer forms an "R-loop" RNA:DNA hybrid by binding to DNA upstream of the heavy strand origin of replication (Lee and Clayton, 1996; Lee and Clayton, 1998; Xu and Clayton, 1996). Due to their position within the D-loop, and the fact that the major initiation site for heavy-strand DNA replication occurs almost exclusively near the first conserved sequence block, it has been hypothesized that the conserved sequence blocks may play a role in stabilizing the R-loop structure which is essential for proper primer binding and subsequent processing (Shadel and Clayton, 1997). RNA processing of the R-loop by the nuclear encoded mitochondrial RNA-processing endoribonuclease leaves a mature RNA primer capable of initiating replication (Chang and Clayton, 1987). The mitochondrial DNA polymerase extends from the RNA primer and begins mtDNA replication.

Following initiation from the heavy strand origin, it is possible that the entire mitochondrial genome may be replicated, but it is more likely that replication proceeds until the mtDNA polymerase encounters the "termination associated sequences" within the D-loop. The termination associated sequences cause replication arrest, leaving the nascent heavy DNA strand to base-pair with the parental light stand, forming the characteristic D-loop structure (Clayton, 1982). It remains unclear what factors determine whether the mtDNA polymerase will terminate or proceed to replicate the entire genome (Brown and Clayton, 2002).

Replication of the light strand relies on the formation of a stable stem-loop structure within the small non-coding light strand origin (Martens and Clayton, 1979; Tapper and Clayton, 1981; Wong and Clayton, 1985a). Likely, the stem-loop structure allows the formation of a replication fork as the mtDNA polymerase passes through the

light strand origin, which exposes the heavy strand to a DNA primase (Shadel and Clayton, 1997). The DNA primase generates short RNA primers, which allow subsequent elongation and replication of the light strand by the mtDNA polymerase (Hixson *et al.*, 1986; Wong and Clayton, 1985b).

Because initiation of replication from the heavy and light strands occurs asynchronously, replication from both origins can occur simultaneously (Fernandez-Silva *et al.*, 2003). Post-replicative modifications that follow complete synthesis of the mitochondrial genome include the removal of RNA primers, ligation of subsequent gaps in sequence, physical separation of nascent molecules from the parental molecules, and adoption of functional tertiary structure (Lecrenier and Foury, 2000).

As the sole mtDNA polymerase, Pol γ possesses both 5' to 3' DNA polymerase and 3' to 5' exonuclease proofreading activity (Johnson *et al.*, 2000). Pol γ is composed of two subunits, a large subunit of approximately 125-140 kDa responsible for its catalytic functions, and a small subunit of approximately 30-54 kDa that likely plays a role in primer recognition and processivity (Fan *et al.*, 1999; Fernandez-Silva *et al.*, 2003; Gray and Wong, 1992). The polymerization rate of Pol γ is quite slow, with replication of the mitochondrial genome occurring in approximately one hour (Graves *et al.*, 1998). A number of accessory factors, such as helicases, topoisomerases, single-stranded binding proteins and DNA ligases are also required for successful mtDNA replication. Many of these factors have been recently described and cloned in humans and/or other organisms (Fernandez-Silva *et al.*, 2003). While the mechanism controlling mtDNA replication remains unknown, a number of factors may regulate replication. In mammalian systems, different tissues have characteristically different amounts of total

cellular mtDNA, depending largely on the cell type's reliance on oxidative phosphorylation as an energy source. Additionally, different mtDNA genotypes co-exist within cells, and their relative abundance may be influenced by external factors, indicating a level of replicative control that may be template dependent (Fernandez-Silva *et al.*, 2003).

Transcription and Translation

Transcription can begin from one of three sites on the mtDNA molecule, all contained within the D-loop regulatory region. The first promoter, responsible for transcription of the light strand, is located at nt 407. Two initiation points exist for the heavy strand, with the majority of heavy strand transcripts originating from the promoter at nt 561 rather than the putative minor promoter at nt 638 (Montoya *et al.*, 1982; Montoya *et al.*, 1983; Taanman, 1999). Despite their close proximity, the heavy and light strand promoters are functionally independent (Chang and Clayton, 1984; Walberg and Clayton, 1983).

Initiation of transcription from the heavy and light strand promoters requires binding of a mitochondrial RNA polymerase, as well as other accessory proteins and transcription factors. Although screening of a human expressed sequence tag database has produced a human candidate gene, the mitochondrial RNA polymerase(s) required for mtDNA transcription has not yet been isolated (Tiranti *et al.*, 1997). Mitochondrial transcription factor A (mtTFA) is a transcription factor that acts with the core mitochondrial RNA polymerase. The 25 kDa transcription factor is very abundant in mitochondria and has domains that are likely involved in DNA binding and recognition (Taanman, 1999). In humans in particular, mtTFA, appears to be required for mtDNA

maintenance. Perhaps playing a regulatory role, nuclear-encoded mtTFA binds to the D-loop and may facilitate an interaction between transcription and replication associated proteins (Ghivizzani *et al.*, 1994). Heterozygous mtTFA mice show a reduction in total cellular mtDNA, and homozygous knockout mtTFA mice are embryonic lethal and completely lack mtDNA, presenting a compelling case for mtTFA involvement in mtDNA replication and maintenance during development (Larsson *et al.*, 1998).

Transcription of a single polycistronic RNA proceeds following initiation at the light strand promoter (Murphy *et al.*, 1975). For the heavy strand, an additional level of regulation is required as two promoters are involved. In order to explain the observation that in rapidly growing cells, rRNAs are synthesized at a much higher rate than the remaining heavy strand coding sequences, it has been hypothesized that the rates of transcription initiation from the two heavy strand promoters are independent. Transcription from the heavy strand promoter at nt 561 begins frequently, and terminates following the 16S rRNA gene, forming a polycistronic RNA encoding both rRNAs. Initiation of transcription from the heavy strand at promoter nt 638 occurs less frequently, but results in a polycistronic RNA that includes the entire heavy strand (Taanman, 1999). Therefore, in rapidly dividing cells, it is likely that transcription is biased in favour of initiation at the heavy strand promoter located at nt 561.

Once polycistronic RNA transcripts are produced, RNA processing occurs to produce mature RNAs. The unique organization of the mitochondrial genome may allow for ease of RNA processing; tRNAs flank both rRNA genes as well as the majority of structural genes, and may signal sites requiring further processing (Ojala *et al.*, 1981). It has been hypothesized that the tRNAs adopt a secondary structure more conducive to

RNA cleavage, which promotes their removal from nascent RNA. mRNAs and rRNAs are then post-transcriptionally polyadenylated, and terminal CCA sequences are added to the 3' end of the newly formed tRNAs (Clayton, 1984).

The process of mitochondrial protein translation consists of three stages: initiation, elongation and termination. Although the process of translation in prokaryotes and within the cytoplasmic compartments have been relatively well characterized, mitochondrial protein translation remains poorly understood. Having little or no 5' untranslated sequences and no 5' 7-methylguanylate cap, mammalian mitochondrial mRNAs do not contain the typical sequences that facilitate ribosome binding (Montoya *et al.*, 1981). The lack of these "ribosomal markers" may relate to the particularly low rate of mRNA translation within mitochondria. In the absence of typical structural cues, translation begins nonetheless at or very near the most 5' methionine residue of the template.

Mitochondrial translation occurs on ribosomes in the mitochondrial matrix and the components of the mitochondrial translational machinery are distinct from their cytoplasmic counterparts. Mitochondrial and cytoplasmic ribosomes are structurally and chemically distinguishable, as mitochondrial ribosomes have particularly low RNA and high protein content (Taanman, 1999). Additionally, mitochondrial protein synthesis occurs in the absence of the 5S rRNA, which is typically present in ribosomes, but the 39S and 28S ribosomal subunits which contain the 16S and 12S rRNA species, are retained within the mitochondrial translational machinery (Attardi and Ojala, 1971; Brega and Vesco, 1971). The majority of the hundreds of proteins required for mitochondrial translation are encoded by the nucleus and must be translated in the cytosol and imported

into the mitochondrial matrix. During translation initiation, the 28S ribosomal subunit interacts with a minimum of 30-80 nucleotides of mRNA; however, efficient binding requires approximately 400 nucleotides (Liao and Spremulli, 1989). The smallest genes (ATPase6 and ATPase8, and ND4L and ND4) are therefore transcribed as dicistronic messages, likely to allow a stronger and more reliable interaction with the 28S ribosome (Liao and Spremulli, 1989). To date, only one other initiation factor has been identified in mammalian systems. mtIF-2 is a GTPase that facilitates binding of Met-tRNA to the ribosome-bound template (Liao and Spremulli, 1991). The elongation factors that have been purified and studied thus far bear a striking similarity to their prokaryotic counterparts. It has therefore been presumed that the process of elongation in mammalian mitochondria occurs in a similar manner to that observed in prokaryotic organisms, as mitochondrial protein synthesis uses very similar elongation factors, and is sensitive to chloramphenicol, which inhibits the activity of bacterial ribosomes (Hammarsund *et al.*, 2001; Wallace, 1982).

Translation of mitochondrial DNA is restricted to the mitochondrial matrix due to differences between the genetic code utilized by the mitochondria and the nucleus. Within the mitochondria, UGA is translated as tryptophan rather than a stop codon, and AGA and AGG serve as stop codons rather than being translated as arginine. As most mitochondrial proteins contain several tryptophan residues, translation of mitochondrial mRNA in the cytoplasm would result in severely truncated proteins (Wallace, 1982).

DNA Mutation Rate and Repair

The rate of mtDNA mutation is elevated 10-20 fold or greater compared to its nuclear counterpart, likely due to a number of additive detrimental features inherent to

the mitochondrion itself. The mitochondrial genome has an extremely high lesion equilibrium frequency, much of which is attributed to damage from hydrogen peroxide and free radicals produced through oxidative phosphorylation (Driggers *et al.*, 1997). Nearly 90% of cellular free radicals are produced in the mitochondria (Shigenaga *et al.*, 1994). Continual exposure to reactive oxygen species leads to an increased amount of oxidative damage, leaving a mutation spectrum highly biased towards transitions (Khrapko *et al.*, 1997). The mtDNA also lacks protective histones, likely an artifact of its prokaryotic ancestry. As such, mtDNA is left exposed to any chemical or mutagen it contacts, contributing to its high observed mutation rate. Furthermore, DNA polymerase γ, the sole polymerase responsible for replication of mtDNA, is known to be error-prone, leading to an increased rate of mispairing and base substitution, despite its 3' to 5' exonuclease activity (Pinz *et al.*, 1995). The proofreading ability of DNA Pol γ appears insufficient to prevent mutation due to oxidative damage, particularly regarding the mispairing between oxidatively damaged guanine and adenine. Until recently, it was believed that mitochondria lacked the means by which to repair the DNA housed within it. Evidence now exists supporting the existence of more primitive mtDNA repair mechanisms, including base excision repair, mismatch repair, direct damage reversal and perhaps recombinational repair (Croteau *et al.*, 1999).

Repair of bulky lesions in the nuclear genome is biased toward prompt repair of transcribed segments of DNA by a process called transcription coupled repair. Within the mitochondria, repair of both oxidative and alkylation damage is repaired by base excision repair. It has been demonstrated that the mitochondrion possesses means of detecting and repairing oxidative damage within its genome. Although there are hotspots

within the mtDNA genome that are particularly susceptible to oxidative damage, repair of mitochondrial damage occurs equally in all regions and does not display a bias toward either the heavy or light strands (Driggers *et al.*, 1997). Additionally, transcribed and non-coding regions are repaired at equal rates, indicating that there is no transcription-coupled repair of oxidative mtDNA damage (Anson *et al.*, 1998).

Base excision repair involves the cleavage of modified bases, such as 8-oxodeoxyguanosine or thymine glycol, by specialized glycosylase enzymes. The mitochondrial uracil DNA glycosylase is an alternatively spliced isoform derived from the same nuclear gene as the nuclear isoform, with a localization sequence at its N-terminus that directs its transport to the mitochondria (Nilsen *et al.*, 1997; Otterlei *et al.*, 1998). Several glycosylases able to specifically cleave thymine glycol have been purified from rat liver mitochondria. Thymine glycol endonuclease shows highly specific glycosylase activity allowing it to correct T/G mismatches, and has inherent apurinic/apyrimic (AP) endonuclease activity.

Mitochondria also have means of detecting oxidized guanine. 8-oxodeoxyguanosine (8-oxodG) is a common adduct formed when guanosine reacts with oxygen free radicals or hydrogen peroxide. 8-oxodG is capable of mispairing with adenine rather than cytosine during DNA replication, and is therefore considered to be a premutagenic lesion (Bohr *et al.*, 2002). A glycosylase that specifically recognizes and incises 8-oxodG when paired with a cytosine has been identified in purified rat liver mitochondria (Croteau *et al.*, 1997). Capable of excising misincorporated adenines in an 8-oxodG:A pair, hMutY adenine DNA glycosylase has also been reported to localize to the mitochondria (Takao *et al.*, 1998). As a third line of defense, mitochondria also

possess an 8-oxodGTPase that degrades free 8-oxodG in order to minimize incorporation into nascent strands of DNA (Kang *et al.*, 1995).

Eukaryotic mismatch repair is able to repair mispaired bases and insertion/deletion loops that occur during DNA replication. Recent reports suggest that mitochondria may harbour primitive mismatch repair capabilities. Single G-G and G-T mismatches were repaired by mitochondrial repair proteins, with the apparent mismatch repair activity occurring independently of the nuclear mismatch repair process. MSH2 nuclear mismatch repair enzyme was not involved in the process of mtDNA mismatch repair (Mason *et al.*, 2003).

The most direct mechanism of preventing a premutation from undergoing subsequent rounds of replication and becoming fixed in the genome is to reverse the DNA damage at the site of the lesion prior to replication. Direct reversal of UV-induced pyrimidine dimers does occur in yeast mitochondria, but it is unknown whether photolyases exist in the mitochondria of higher eukaryotes (Croteau *et al.*, 1999). Removal of alkyl or methyl adducts can be performed by DNA alkyltransferases, which redistribute methyl or alkyl adducts from DNA bases directly to a cysteine residue within the repair enzyme, thereby rendering the enzyme inactive. O^6-methylguanine-DNA methyltransferase removes methyl groups from guanine residues and has been partially purified in rat liver mitochondria (Myers *et al.*, 1988)

Mitochondria likely possess a DNA double strand break repair activity similar to that involved in homologous recombination or end-joining pathways in the nucleus (Lakshmipathy and Campbell, 1999). Mitochondrial protein extracts required ATP and magnesium to join the ends of both cohesive and blunt-ended DNA fragments. In cases

where the end-joining process was erroneous, the patterns of deletions observed bore a striking resemblance to the deletions commonly implicated in Kearns-Sayre or Pearson syndrome, both neurodegenerative disorders caused by mutations in the mtDNA genome.

Recombinational repair occurs extensively in yeast, but whether mammalian mtDNA undergoes recombination remains a subject of debate (Dujon *et al.*, 1974). The existence of a means of re-joining double-stranded breaks, as well as evidence of mtDNA complementation between fused human mitochondria, adds support to the belief that inter-strand recombination may be a means of correcting mtDNA errors (Lakshmipathy and Campbell, 1999; Ono *et al.*, 2001).

Mitochondrial Genetics

Maternal Inheritance

The genetics involved in the study of mitochondrial inheritance differ radically from Mendel's accepted laws of inheritance. The first unique feature of mtDNA genetics is maternal inheritance, meaning that the mother, but not the father, will pass her mtDNA to all of her offspring. Maternal inheritance is likely the consequence of the number of molecules of mtDNA that inhabit the ovum and sperm; the mammalian ovum contains approximately 100,000 molecules of mtDNA whereas the sperm contains on the order of hundreds of molecules of mtDNA (Chen *et al.*, 1995; Manfredi *et al.*, 1997). Although mtDNA derived from the sperm is permitted to enter the cytoplasm of the ovum upon fertilization, and paternal mtDNA can be detected at the four- and eight-cell stage of certain embryos, paternal mtDNA cannot be detected in the later stages of embryogenesis (St John *et al.*, 2000). Elimination of sperm mtDNA during embryogenesis may occur by selective degradation or inactivation, or the sperm mtDNA simply be outnumbered by the

number of maternal mtDNA molecules in the embryo, so that at later stages of embryogenesis, the paternal mtDNA is lost completely due to dilution (Shitara *et al.*, 2000; Sutovsky *et al.*, 1999).

Recently, a report of paternal inheritance has questioned the doctrine of exclusive maternal inheritance of mtDNA. The case of paternal inheritance, though admittedly rare, suggests that mitochondrial DNA from the sperm may under certain conditions survive the process of degradation or inactivation observed during embryogenesis. Furthermore, the paternally inherited mtDNA was present in approximately 90 percent of the patient's skeletal muscle mtDNA, indicating that the relatively small mtDNA contribution from the sperm mitochondria managed to survive vast dilution from ovum derived mtDNA to comprise the major mtDNA genotype within the affected tissue (Schwartz and Vissing, 2002).

Homoplasmy and Heteroplasmy

Mitochondrial genetics are governed largely by the number of DNA molecules per cell. In contrast to the diploid nature of the nuclear DNA, the number and content of mtDNA molecules varies between different cell types within an individual (Lightowlers *et al.*, 1997). Most somatic mammalian cells contain thousands of mitochondrial DNA molecules housed within hundreds of mitochondria (Bogenhagen and Clayton, 1974; Nass, 1969). Each mitochondrion contains approximately two to ten molecules of mtDNA (Satoh and Kuroiwa, 1991). Given that there are a vast number of mtDNA molecules per cell, each cell can house identical mtDNA genotypes (termed homoplasmy) or a mixture of different mtDNA genotypes (termed heteroplasmy).

Heteroplasmy within the coding region of the mtDNA is typically associated with mitochondrial disease. As the mtDNA lacks introns and intergenic spaces, nearly any mutation that occurs will affect a coding region (Johns, 1995). A "threshold effect" governs mitochondrial disorders; the phenotype of the disorder appears only once a minimum mutational threshold has been reached, and the relative abundance of mutant genomes in affected tissues determines the severity of the observed phenotype (Kirches *et al.*, 2001; Wallace, 1992a). For example, a mutation at nt 8993 with a frequency of less than 75% results in neurogenic weakness, ataxia, and retinitis pigmentosum (NARP), while levels above 95% leads to early onset Leigh syndrome, which results in ataxia, hypotonia, spasticity, developmental delay and opthalmoplegia, and is often lethal (Wallace, 1999). The threshold level of mutant mtDNAs required for a phenotype depends both on the mitochondrial disorder as well as the energetic needs of the tissue involved. Tissues with the highest energy requirements show the highest sensitivity to mutant mtDNA levels. As the level of mtDNA mutation increases, the ability of the mitochondria to produce ATP gradually decreases, until eventually they are unable to meet the energetic levels required for normal tissue and organ function (Wallace, 1992b).

In order to determine the frequency of cellular heteroplasmy, Monnat *et al.* utilized recombinant and enzymatic techniques to determine the level of heteroplasmy in human neoplastic cells (Monnat *et al.*, 1985). Upon observation of nearly 400 clones, it was determined that heteroplasmy could be detected only very rarely within this cell type. From this observation, it has therefore become generally accepted that normal individuals are homoplasmic, containing a single mitochondrial genotype. This observation has recently been brought into question with the study of different tissue types in single

individuals. Initially stemming from a report of individuals with heteroplasmic mtDNA polymorphisms within the D-loop region, investigation into different populations and cell types has produced evidence that the level of heteroplasmy in the normal population may in fact be higher than originally thought (Comas *et al.*, 1995). Denaturing gradient gel electrophoresis and sequencing results from different tissues demonstrate that although no significant incidences of heteroplasmy could be detected in leukocytes, normal brain tissue exhibited very high levels of heteroplasmy, with up to 45% of the molecules examined displaying one or more sequence changes distinct from the remaining tissue. Moreover, the heteroplasmy identified in the brain tissue was restricted to the non-coding region of the mtDNA, as no heteroplasmy could be detected in the coding region (Jazin *et al.*, 1996). This study, as well as several reports investigating the level of heteroplasmy in other species, has suggested that the level of human heteroplasmy may have been previously underestimated (Lightowlers *et al.*, 1997).

Based on the observation that normal individuals may be heteroplasmic, a further study was performed by Kirches *et al.* examining the proportion of normal and mutant mtDNAs within tissues of a single individual using a single polymorphic marker (Kirches *et al.*, 2001). Using a frequently expanded tract of seven consecutive cytosines located at nt 303 (C-tract) as a marker for heteroplasmy, mtDNA from normal liver, lung, heart, kidney, brain, muscle and spleen were examined for organ-specific mtDNA polymorphisms. The study demonstrated that heteroplasmy was variable in terms of tissue distribution and fluctuations in C-tract length were observed in most organs examined. Despite their shared origin, it appears that during development, different organs develop distinct populations of mitochondrial genomes. The distribution of C-

tract lengths was not specific to any given organ, but instead depended solely on the individual. For example, in one patient, the C-tract was shorter in the kidney and longer in the liver and the lung, whereas in two other patients, the situation was exactly reversed. The observation of multiple populations within a given tissue type in an individual, as well as the indefinite nature of the C-tract distribution between organs in different individuals has led to the suggestion that there exists a strong imbalance in the mtDNA composition among organs. The random nature of mtDNA distribution during embryogenesis also leads to large mtDNA differences between individuals (Kirches *et al.*, 2001).

In addition to tissue-specific variation in mtDNA heteroplasmy, the existence of temporal-specific variation may also exist within individuals. A study of human fibroblast cells from both young and aged subjects revealed a significant increase in point mutations, and hence heteroplasmy, in subjects over 65 years of age. Longitudinal studies also demonstrated the appearance of mtDNA polymorphisms over time as the subject aged (Michikawa *et al.*, 1999). This observation was corroborated by a second group examining mtDNA from blood, heart, muscle and brain tissue from 43 individuals of different ages. Not only did the level of heteroplasmy vary with tissue type, as would be expected from the study by Kirches *et al.* (Kirches *et al.*, 2001), but older individuals displayed a statistically significant increase in heteroplasmy when compared to younger individuals, indicating that mtDNA mutations and polymorphisms likely accumulate with age (Calloway *et al.*, 2000).

In contrast to the publications of Michikawa *et al.* (Michikawa *et al.*, 1999) and Calloway *et al.* (Calloway *et al.*, 2000), reports have emerged showing that heteroplasmy

remains constant with age. For example, a longitudinal study examining cervical cell samples seems to suggest that heteroplasmy is not the result of somatic age-related events (Lagerstrom-Fermer *et al.*, 2001). Samples displaying heteroplasmy at polymorphic markers located at either nt 309 or nt 16189 were assayed by solid-phase minisequencing to determine the level of heteroplasmy over a span of one to two decades. This study demonstrated that the heteroplasmic levels of these two polymorphisms remained unchanged over the course of the study. However, as this study only examined the level of heteroplasmy at two individual sites within the mtDNA genome, it is unclear whether heteroplasmy at other sites, especially within the non-coding regulatory region, may vary with time. Also, since only a single cell type was analyzed, results may not be extrapolated regarding other cell types within the individuals studied.

Evidence for site and tissue-specific age-related increases in heteroplasmy has received support from a study of the frequency of a polymorphism at nt 414, and mutations at nt 3243 and nt 8344 in individuals aged 23-93 (Murdock *et al.*, 2000). While no age related mutations appeared in either skeletal muscle or brain tissue at nt 3243 and nt 8344, polymorphisms at nt 414 accumulated with age in the skeletal muscle, but not in the brain tissue. This research, in conjunction with other studies reported in the literature, provides a case for tissue- and temporal-dependent heteroplasmy in individuals but no definitive formula has been derived to determine which individuals will display heteroplasmy and which will not.

The mechanism by which heteroplasmy develops and is transmitted remains open to debate. It is unclear whether the tissue specific nature of heteroplasmy is driven by mitochondrial genotypic selection within the tissue or by random drift alone. If

mutations occur in mtDNA coding regions, they have the potential to alter the level of oxidative phosphorylation in cells and may therefore improve their fitness for their environment. For example, a study of mtDNA mutations in skeletal muscle demonstrated that a small proportion of mutant mtDNAs may have a dominant effect in cases where mitochondrial myopathy, and hence cytochrome oxidase deficiency, was present. Despite a majority of wildtype mtDNA molecules, the mutant mtDNAs had the ability to affect the level of heteroplasmy within the skeletal muscle (Shoubridge *et al.*, 1990). It is intuitive to assume that mutations that increase the fitness of an organism may undergo positive selection.

Alternatively, mutations or polymorphisms within the coding region of the mtDNA may affect the efficiency of DNA replication. It is quite possible that a mutation could cause either positive or negative selection of mutant mtDNA populations. For example, the proportion of genotypes containing the pathogenic MELAS A3243G mutation decreases in blood over time, as would be expected of a harmful mutation if negative selection were in place (t Hart *et al.*, 1996). In contrast, the same mutation in cell lines has been shown to confer an intracellular replicative advantage, significantly above wildtype levels, to the MELAS-containing mutant mtDNA genotype (Yoneda *et al.*, 1992).

It has also been suggested that shorter mtDNA genomes may confer a replicative advantage over longer genomes, as mtDNA replication is asynchronous and mtDNA polymerase γ may replicate shorter molecules more quickly, hence affecting the heteroplasmic balance within cells (Moraes *et al.*, 1999). Shorter mitochondrial genomes were able to repopulate a mtDNA-less cell line more rapidly than longer genomes.

However, Tang *et al.* found that in steady states, deleted mtDNA genomes were not selected over full-sized wildtype genomes in long term cultured cell lines (Tang *et al.*, 2000)

Genetic studies of Holstein cows have revealed that the distribution of heteroplasmic polymorphisms in the offspring is often altered relative to the maternal distribution, with the polymorphism at times becoming homoplasmic in successive generations (Olivo *et al.*, 1983). Similar situations have since been reported in human pedigrees, where in some cases, certain polymorphisms appeared homoplasmic after a single generation (Olivo *et al.*, 1983). While random genetic drift can easily explain slight variations in the degree of heteroplasmy from one generation to the next, a unique aspect of mitochondrial genetics in which genomes are not equally distributed to offspring during cytokinesis, called vegetative segregation, may be required to help explain more dramatic shifts (Chinnery *et al.*, 2000).

Vegetative segregation resulting in unequal partitioning of genomes occurs both during mitosis and during meiosis. During mitosis, mitotic segregation occurs over repeated cell divisions, causing the relative levels of wildtype and mutant mitochondrial genomes to drift, leading to the observed mtDNA variation between cell types. During meiosis, if meiotic segregation occurs resulting in random, unequal partitioning of a large number of genomes, a typical pattern of random segregation would likely result in the subsequent offspring.

It is more difficult to explain how a polymorphism may become fixed within a single transmission. Mathematically, it is not impossible to explain dramatic shifts using the paradigm of random genetic drift alone (Coller *et al.*, 2001). However, if the

effective number of mtDNA molecules involved in reproduction was reduced, resulting in a so-called "mitochondrial genetic bottleneck", only a small group of mtDNA molecules would genetically contribute to the following generation. The potential for a great genetic shift increases dramatically as the sample size of the repopulating group decreases, allowing the level of heteroplasmy to shift at a much greater rate (Chinnery, 2002). In mice, it has been demonstrated that a bottleneck does indeed occur early in oogenesis, and that the number of mtDNA molecules participating in the process of repopulation is approximately 200, meaning only 0.2% of the possible mitochondrial genomes are available to future generations (Jenuth *et al.*, 1996).

Mitochondrial DNA Mutation and Disease

The phenotypic presentations of most mitochondrial diseases are very diverse. Often, a single mutation can result in very different phenotypes in different patients, depending on the tissue affected and the level of heteroplasmy within the given tissue. Furthermore, mutations at different sites of the mtDNA can cause similar clinical phenotypes. It is often difficult for physicians to determine the nature of the mtDNA disorder through observation of patient symptoms alone. As a result, mitochondrial diseases are often classified by their underlying mitochondrial mutation rather than the phenotypes the mutations invoke (Wallace, 1999)

MtDNA mutations fall into one of two classes: single base substitutions and rearrangement mutations (Wallace, 1999). Mutations can be subdivided further by classifying the effect of the mutation, with certain mutations affecting mitochondrial protein synthesis and others affecting the protein coding genes directly (DiMauro and Schon, 2001). Mutations affecting mitochondrial protein synthesis most often involve

deletions of one or more of the tRNA genes, or point mutations in either tRNA or rRNA genes, thus affecting the ability of the mitochondria to translate mRNA into protein (DiMauro and Schon, 2001). Mutations of this type are often associated with multi-systemic symptoms, as the translation of all protein coding genes of the mtDNA is affected by the reduced level of mtDNA translation, although in rare circumstances a single tissue is affected, most commonly the skeletal muscle. The two most common mitochondrial protein translation disorders are MELAS (mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes) and MERRF (myoclonic epilepsy with ragged red fibers). MELAS is most commonly caused by a mutation in the tRNALeu(UUR) gene, with onset occurring in childhood or early adolescence. Approximately 80% of MELAS cases are attributed to a mutation at nt 3243, a site that has also been implicated in other clinical phenotypes, including myopathy, deafness, chronic progressive external ophthalmoplegia (CPEO), and maternally inherited diabetes mellitus (Simon and Johns, 1999). MERRF symptoms include seizures, ataxia and mitochondrial myopathy and usually occur during early adulthood (Leonard and Schapira, 2000). MERRF is caused by a mtDNA mutation in the tRNALys gene, with mutations occurring most commonly at nt 8344, and less commonly at nt 8356 and nt 8363 (Simon and Johns, 1999)

Mutations affecting mitochondrial protein coding genes vary immensely in their phenotypes. LHON (Leber's hereditary optic neuropathy) results in sudden, painless bilateral vision loss and optic atrophy, with onset typically occurring in adulthood. It is caused most often by one of three mutations, occurring at nt 3460, nt 11778, or nt 14484, which fall within the ND1, ND4, and ND6 protein coding genes, respectively (Simon and

Johns, 1999). LHON-associated mutations tend to be homoplasmic and most family members that carry mutations in the heteroplasmic state are asymptomatic. Males are nearly four times as likely as females to develop LHON, likely due to differences in the nuclear genetic background (Johns, 1995).

Patients harbouring a single point mutation at nt 8993 of the ATPase 6 gene suffer a much more severe phenotype than LHON, with a marked decrease in ATP synthesis due to malfunction of the ATPase protein. Individuals carrying a missense mutation at nt 8993 at heteroplasmic levels below 75%, develop NARP (neuropathy, ataxia, and retinitis pigmentosa); however, when the mutation load exceeds 95%, the patient develops early onset Leigh's syndrome, which is often lethal (Wallace, 1999)

Cancer: Endometrial Cancer and Multiple Myeloma

Endometrial Cancer

Nearly 140,000 new cases of cancer and 67,400 deaths will occur in Canada in 2003. Endometrial cancer is the fourth most common cause of newly reported cancer cases in Canadian women. Endometrial cancer is very treatable if detected early, yet it remains the ninth most common cause of cancer-related death in women (National Cancer Institute of Canada, 2003). These statistics indicate that a method of early detection is essential in preventing unnecessary cancer-related deaths.

Endometrial cancer appears to develop either as a result of hormonal imbalances, and their subsequent effects, or independently of hormonal action, although the molecular and cellular events that drive endometrial carcinogenesis remain largely unknown (Sherman, 2000). Risk factors for endometrial cancer include obesity, type II diabetes and hypertension. The use of unopposed estrogens in hormone replacement therapy in

women increases their risk for endometrial cancer six-fold. Administration of Tamoxifen, an estrogen blocking drug used in the prevention or treatment of breast cancer, may double or triple a woman's risk of developing endometrial cancer. Although its frequency in the general population is relatively low, there is a significant genetic risk of developing endometrial cancer in individuals predisposed to hereditary non-polyposis colorectal cancer, a genetic condition resulting from a deficiency in the nuclear mismatch repair machinery (Robertson, 2003). The lack of mismatch repair is reflected in the expansion or contraction of short sequences of repetitive DNA motifs called microsatellite sequences. Genetic microsatellite instability is a hallmark of cells lacking mismatch repair, and several markers have been developed to distinguish between mismatch repair proficient and deficient cancer types.

Multiple Myeloma

In 2003, approximately 1800 new Canadian cases of multiple myeloma will be diagnosed, with all cases proving fatal. Indeed, multiple myeloma has the fifth highest mortality rate of cancers in Canada, falling just below esophageal, stomach, pancreatic and lung cancers (National Cancer Institute of Canada, 2003). Multiple myeloma is a hematological cancer caused by continuously replicating B lymphocytes, which is initially restricted to the bone marrow, but may eventually spread outside of the bone marrow, especially to the blood and skin (Kuehl and Bergsagel, 2002). Believed to follow a multi-step process, multiple myeloma is characterized by karyotypic changes, with most tumours exhibiting translocations involving the immunoglobulin locus (Bergsagel *et al.*, 1996). Although the associated site of translocation is often varied, there appears to be some bias toward translocations between the immunoglobulin locus

and certain cyclins. Nearly 25% of translocations involve cyclin D1, D3 and possibly D2, while the remaining translocations appear to exploit genes exhibiting other oncogenic effects. The c-MYC and K-RAS oncogenes becomes dysregulated in late stage multiple myelomas, and this appears to correlate with enhanced cell proliferation (Kuehl and Bergsagel, 2002).

Treatment of multiple myeloma has historically been restricted to non-specific anti-inflammatory and alkylating drugs, however, many biologically-based therapies are currently under investigation. Patient survival varies significantly with the nature of the translocation present; it is hoped that mutation-specific therapies will be forthcoming in treating patients harbouring the 'characteristic' genetic aberrations involving the immunoglobulin locus and various cell cycle and oncogenic proteins (Kuehl and Bergsagel, 2002). A marker that would relay information regarding the stage, and perhaps prognosis or responsiveness to certain therapies, would greatly reduce the morbidity and mortality associated with multiple myeloma.

Mitochondrial DNA and Cancer

Mutations in the nuclear genome and their contribution to the multi-step cancer development process have been studied quite extensively, but until recently the contribution of the mitochondrial genome in this process has been largely overlooked. Mitochondrial defects have long been implicated in cancer development. Hypotheses dating back over 50 years attempt to explain the observed shift in energy metabolism in malignant cells when compared to their normal counterparts (Warburg, 1956). The so-called "Warburg effect" is used to describe the reliance of cancer cells on glycolysis, leading to the production of high levels of lactate and pyruvate (Lu *et al.*, 2002). The

relevance of the Warburg effect in the process of carcinogenesis is unclear, however, it has been observed that increased reliance on glycolysis appears to be correlated with cancer metastasis in both cervical and head and neck cancers (Schwickert *et al.*, 1995; Walenta *et al.*, 1997). Changes in mitochondrial physiology, therefore, appear to have a significant impact on cancer development and prognosis.

Much attention has recently been given to the ability of mitochondria to affect the balance between prevention and promotion of cancer through its role in the initiation of apoptosis. Apoptosis is initiated through one of two pathways: the extrinsic pathway involves cues transmitted via death receptors on the cell membrane, and the intrinsic pathway involves the mitochondria. The mitochondrial pathway is regulated by Bcl-2 family members, which can be pro-apoptotic or anti-apoptotic. Following apoptotic activation, cytochrome *c* and apoptosis inducing factor are released from the mitochondria into the cytosol, and the mitochondrial membrane potential drops. Intracellular proteases, caspases, are then activated which propagate the characteristic cellular and morphological features of apoptosis, including chromatin condensation, membrane blebbing and cell fragmentation. Faulty apoptotic mechanisms, which are inherent to some cancer cells, allow tumours to increase in size due to loss of contact inhibition, which often interferes with cancer therapies that attempt to trigger apoptosis (Igney and Krammer, 2002). In addition, it has been demonstrated recently that levels of apoptosis increase in cells lacking either mitochondrial transcription factor A or mtDNA (Wang *et al.*, 2001). Disruption of mtDNA expression, and the resultant respiratory deficiency leading to increased apoptosis, may thus be a major pathogenic factor in mitochondrial disorders. In addition, the authors suggest that this observation could be

manipulated in cancer cells to increase their responsiveness to chemotherapy and radiation treatments.

Not only do mitochondria affect the potential tumorigenicity of a cell via apoptosis, but evidence also suggests that mitochondrial oxidative or genetic stress can promote carcinogenesis and invasive phenotypes, suggesting a potential mechanistic link between mitochondrial dysfunction and cancer development (Amuthan *et al.*, 2001; Bandy and Davison, 1990). Disruption of the mitochondrial membrane or depletion of mtDNA caused up-regulation of certain nuclear genes involved in tumorigenesis, including cathepsin L, a lysosomal cysteine protease commonly secreted by malignant cells, and transforming growth factor (TGFß), a tumour suppressor involved in cellular growth, differentiation and tumorigenesis. When grown in culture, normally non-invasive myoblast and pulmonary carcinoma cells demonstrated a four- to six-fold increase in invasiveness. Surprisingly, when the wildtype mtDNA populations in the cells are reverted to near-normal levels (approximately 70%), the invasiveness of the cultured cells returned to baseline levels, indicating that mitochondrially-induced tumorigenic effects are essentially reversible in these cells.

Since the first report in 1998, a trend towards examining mtDNA for tumour-specific variations has developed (Polyak *et al.*, 1998). In this initial study, the complete mtDNA genome was PCR amplified and sequenced in order to detect mutations and/or polymorphisms in ten human colorectal cancer cell lines. When the sequences of the colorectal cell lines were compared to the mitochondrial reference sequence, 88 sequence variants were obtained, with individual cell lines displaying between four and 31 variants. In order to determine which mutations represented germline variants and which were

acquired during the process of tumourigenesis, the tumour sequence was compared to sequence data obtained from normal colon tissue from the same patient. This further analysis reduced the number of somatic mutations to 12, with three of the cell lines containing a single mutation, one cell line containing two mutations, and three cell lines containing three mutations. The mutation spectrum was biased toward C to T transitions, consistent with spontaneous deamination of cytosine, and G to A transversions, consistent with oxidative damage to guanine. With the exception of a single deletion, all somatic changes were single base pair substitutions, while no large scale deletions were detected in this study. Interestingly, all mutations occurred in protein coding or rRNA genes, and no mutations or polymorphisms were reported in the non-coding D-loop regulatory region, which encompasses the two hypervariable regions which are known to be mutational hotspots. The origin of the mtDNA mutations in this study remains unknown; whether the mutant mtDNA genomes became fixed in the tumour population by chance or the mutant genome plays a role in the carcinogenic process has yet to be fully investigated.

The observation of somatic mutations in tumour cells was coupled to the more surprising observation that the majority of the somatic mutations were homoplasmic within tumour cells (Polyak *et al.*, 1998). The authors suggest that the presence of homoplasmy indicated some replicative advantage inherent to the mutant mtDNA molecules and selection by mitochondria that harbour the mutant mtDNAs. Coincidentally, the presence of the mutant mtDNA molecules must have occurred in a cell that had acquired a selective growth advantage during the evolution of the tumour. The authors suggest that the presence of homoplasmic, tumour-specific mutations may

provide a useful tool in cancer diagnosis, and that microarray technology may allow rapid, cost effective cancer-related mtDNA screening.

Further studies of the mtDNA genome in cancer have evaluated the location and level of heteroplasmy of mtDNA mutations in a wide variety of cancers (Carew and Huang, 2002; Penta *et al.*, 2001). One sequence variant in particular, a cytosine-tract (C-tract) length polymorphism at nt 303, has been hailed as an excellent candidate marker for cancer detection (Sanchez-Cespedes *et al.*, 2001). The C-tract is a mononucleotide repeat sequence, typically containing seven consecutive cytosine residues, followed by a thymine and five cytosine residues (CCCCCCCTCCCCC). Although the reference sequence reports seven cytosines as the normal state of the mtDNA, normal individuals often display a length polymorphism at this site, and the cytosine tract can readily expand to eight or nine consecutive cytosine residues. The C-tract length was determined in a total of 247 primary tumours and compared to the C-tract length of matched lymphocyte DNA from the same patient. In 22% of cases, somatic variations (insertions or deletions) were detected in the tumour compared to the normal germline sequence. Somatic variations were detected in lung, gastric, breast, colorectal, and bladder tumours, but not detected in ovarian or prostate tumours. Importantly, when subclones of the normal lymphocyte samples were sequenced, clones could be detected that displayed the "tumour specific" C-tract profile. Despite this observation, the authors simply offer a cautionary note, explaining that the results are insignificant because the Poisson distribution of the clonal populations was different in the tumours than in the lymphocytes. Efforts to promote this region as a diagnostic marker continue, despite the relatively low prevalence of somatic changes in tumours when compared to normal tissues, and detectable levels of

these "tumour specific" changes in the normal tissues themselves. Nevertheless, several publications have reported somatic length polymorphism at this locus (Ha *et al.*, 2002; Parrella *et al.*, 2003; Parrella *et al.*, 2001).

In order to determine whether mtDNA mutations or polymorphisms may serve as a marker for endometrial carcinoma, Liu *et al.* have recently investigated the D-loop, 12S and 16S rRNA gene sequences in 50 primary endometrial carcinomas and their adjacent normal tissues, including cervical tissue and/or lymphocytes (Liu *et al.*, 2003b). Direct sequencing of PCR amplified DNA was performed in order to determine sites of mutation and their frequency. 28 of 50 endometrial tumours possessed one or more mutations, including point mutations, deletions, and expansions or contractions of mitochondrial microsatellite regions, that were not present in normal tissue from the individual. One large deletion (approximately 50 bp) and seven point mutations were detected, with the deletion and three of the seven point mutations occurring within the D-loop. Microsatellite regions within the D-loop were different in endometrial tumour samples and their corresponding normal controls in 50% of cases. In order to explain the dramatic level of mitochondrial microsatellite instability, the authors suggest differential means of repair as the causative factor, despite recent evidence for mismatch-like repair of mismatched DNA loops in mtDNA, and suggest that random genetic drift could satisfactorily explain the genetic differences accrued during carcinogenesis (Mason *et al.*, 2003). Microsatellite instability, especially in the C-tract at nt 303-309, has been suggested as a potential new tool for early endometrial cancer detection.

Three markers within the D-loop region were analyzed for associated susceptibility to certain types of cancers (Liu *et al.*, 2003a). The frequency of T16189C,

T16519C and a CA-repeated sequence at nt 514, all common polymorphisms, were determined in 51, 67, and 53 samples of primary endometrial, cervical and ovarian carcinomas, respectively. According to the authors, of the three markers, T16189C appeared to be associated with endometrial cancer (43.1% association), but not with normal controls (21.5% association). The frequency of the T16189C polymorphism in endometrial cancer was higher than the frequency in ovarian cancer ($p = 0.045$) and cervical cancer ($p = 0.064$). Despite claims of statistical significance and substantial association between the T16189C marker and endometrial cancer, it appears that any association between mtDNA polymorphisms and endometrial cancer requires further study before a noteworthy candidate for early detection becomes available.

As mtDNA is evidently varied in tumours compared to other tissues in the body, the question has arisen regarding the effect of chemotherapy and radiation on mtDNA. Given that cancer treatments often concentrate on the introduction of DNA lesions in the hopes of inducing apoptosis or necrotic cell death, it is plausible that the state of the mtDNA may mirror the efficacy or responsiveness of the patient undergoing treatment. As a whole, patients who have undergone cancer treatment display a significantly increased level of heteroplasmic mtDNA point mutations compared to untreated patients (Carew *et al.*, 2003; Wardell *et al.*, 2003). It is imperative to determine whether the mutations in the mitochondrial genome are a driving force in the development of carcinogenesis; if mtDNA mutations are in some way promoting the development or progression of cancer, inducing further mutations through cancer treatment could be counterproductive. Furthermore, it was demonstrated by Carew *et al.* that the increased mtDNA mutation level appears to correlate to increased generation of reactive oxygen

species (Carew *et al.*, 2003). The effect of increased production of reactive oxygen species has not yet been fully elucidated in the cell. Generally, the increased level of mutagenic agents should increase the overall genetic instability of the cell. Conversely, if cancer cells display an exceptionally high level of reactive oxygen species, they may provide a doorway for new cancer drugs and therapeutics.

Finally, mtDNA would make an ideal marker for cancer detection as its copy number is thousands of times greater than that of the nuclear genome, allowing much more sensitive detection. This fact has enabled the detection of mtDNA mutations in bodily fluids of cancer patients (Fliss *et al.*, 2000). Homoplasmic mtDNA mutations originating from tumour cells in bladder, head and neck, and lung cancer were readily detectable in urine, saliva, and bronchoalveolar lavage fluids, respectively. Furthermore, the mutated mtDNA was over 200 times more abundant than the nuclear marker, mutated p53. Additional evidence suggests that it will likely be possible to extract mtDNA from blood plasma (Anker and Stroun, 2001). If a diagnostic mtDNA marker were to become available, testing of a single blood sample may be able to detect cancers in their earliest stage of development. The ability to reliably detect tumour-associated mtDNA mutations from more readily available non-tumour tissues may provide a mechanism for sensitive, noninvasive cancer detection. Further advances in cancer diagnostics and the study of mtDNA mutations will likely take advantage of newly emerging nanotechnologies.

Microfluidic Chip Technology

Currently, studying genetic developments involved in cancer development and progression involves costly and time consuming biological methods, comprised primarily of amplification and sequencing of heterogeneous tumour cell populations. The

development of a microfluidic chip-based system for the genetic analysis of mtDNA from bodily fluids and single cells would be greatly beneficial to cancer research. This technology enables integration of many labour-intensive processes, such as DNA amplification, mutation analysis and separation, on a credit card-sized chip. Such "microlaboratories" are potentially very efficient and cost-effective, and would allow the manipulations necessary to study cancer development and progression and give rise to enhanced cancer screening techniques.

A simple cross-structure microfluidic chip (Figure 1-2) can be used to manipulate mtDNA and analytical reagents by applying electric fields, then detecting genetic aberrations by using a combination of DNA intercalating dyes and laser induced fluorescence. DNA, originally loaded in the sample well (CB0), travels into the injection channel as a result of a positive applied voltage to the sample waste well (CB2). Once the stream of DNA reaches the intersection between the injection and separation channels, a plug of DNA may be injected down the separation channel by interchanging the voltage connections so that the electrode immersed in the buffer waste well (CB3) is now positive, attracting the negatively charged DNA. As the entire microfluidic chip is initially filled with a sieving matrix containing a DNA-binding fluorophore, the size of the DNA being tested, as well as possible mutations may be detected using a laser-induced fluorescence system. DNA flows through the sieving matrix as it would an agarose gel, with the larger or more sterically hindered molecules moving more slowly toward the positive electrode than the smaller or less hindered molecules.

Single manipulations such as on-chip DNA sizing, concentration, amplification, enzymatic manipulations and detection have been the focus of most publications;

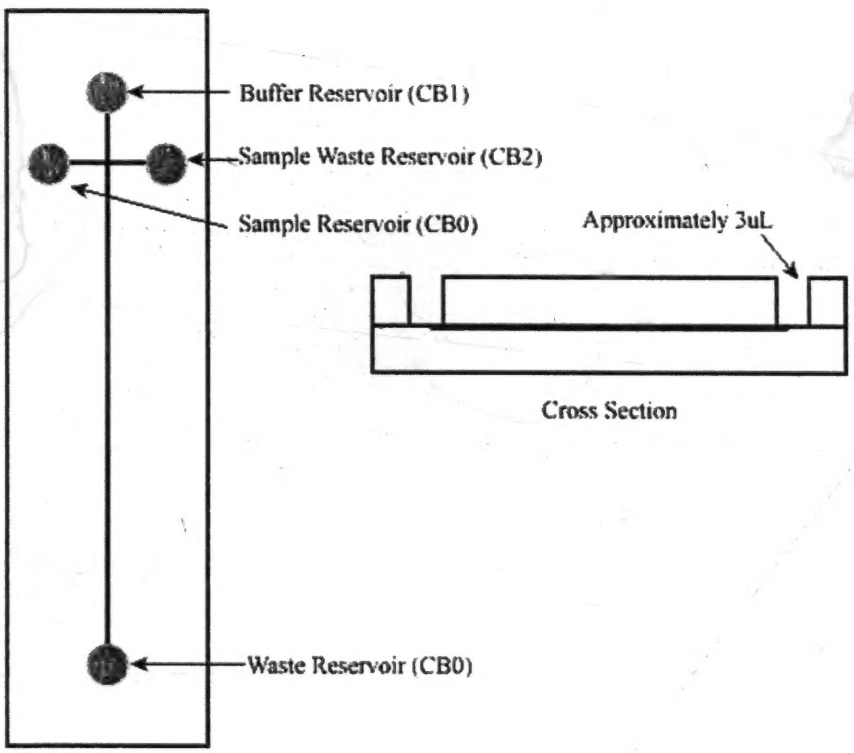

Figure 1-2. Simple cross microfluidic chip design. The microfluidic chips utilized for restriction digestion and heteroduplex analysis were approximately 10 cm in length, and consisted of four sample wells (sample, sample waste, buffer and buffer waste), and two microchannels (a short injection channel and a long separation channel). Sytox Orange intercalating dye was used for double stranded DNA imaging at 532 nm, 76 mm from the intersection between the injection and separation channels.

however, focus is quickly shifting toward more integrated manipulations (Backhouse *et al.*, 2003; Bruin, 2000; Footz *et al.*, 2001; Obeid *et al.*, 2003). In terms of cancer research, rapid mutation detection using heteroduplex analysis or single strand conformation polymorphism (SSCP), which detect sequence changes in double stranded and single stranded DNA, respectively, would benefit greatly through integration with microfluidic technology. The microfluidic chip platform is readily adaptable to the study of mtDNA polymorphisms, and successful integration of enzyme digestion and mutation detection would represent a significant advance in the microfluidic field and its application to DNA diagnostics.

Summary

Mitochondria are known to play a role in carcinogenesis and recent literature suggests an association between the presence of non-inherited changes in the mitochondrial DNA of tumours and cancer progression. The expansion of a poly-cytosine tract at position 303 in hypervariable region II and the propensity for non-inherited changes to be homoplasmic have been proposed as carcinogenic markers. In this thesis, I have employed a subcloning-based sequencing approach to investigate the presence of these markers with a level of sensitivity not yet reported in the literature. I determined the mtDNA sequence of the two hypervariable regions in 16 endometrial tumors, 8 multiple myeloma, and 19 normal blood control samples. These results illustrate that normal individuals display a high level of polymorphism, C-tract length variance, and heteroplasmy. There were no mtDNA changes in either the multiple myeloma or endometrial carcinoma samples that distinguished them from normal samples. The results demonstrate the prevalence of C303 tract length polymorphism and

heteroplasmy in normal as well as cancer samples and suggest that these are not appropriate markers for cancer diagnostics. The implementation of a microfluidic chip based assay for mtDNA has allowed us to integrate restriction enzyme digestion, DNA labeling, and heteroduplex analysis for the first time, resulting in a highly sensitive means of detecting mtDNA variation.

Chapter 2: Materials and Methods

Mitochondrial DNA Specimens

MtDNA from 16 endometrial tumors, eight multiple myeloma samples, and 19 normal blood controls was examined. The endometrial tumours studied included five matched pairs of endometrial tumor and normal blood samples (Dr. Paul Goodfellow, Department of Surgery, Washington University School of Medicine), and 11 fresh or paraffin-embedded endometrial tumors (Dr. Susan Andrew, Department of Medical Genetics, University of Alberta). All endometrial samples were previously analyzed for microsatellite instability in the nucleus and grade of tumor (Table 2-1). Likewise, the eight multiple myeloma cancer samples (Dr. Linda Pilarski, Department of Oncology, University of Alberta) was previously analyzed in terms of cancer grade and patient treatment (Table 2-2). DNA from two normal fibroblast cell lines and 12 normal blood samples (Dr. Diane Cox, Department of Medical Genetics, University of Alberta) was used to obtain control sequences. Although the precise ages of the normal controls are unknown, it can be assumed that they vary between 20-50 years of age. All samples were collected with informed patient consent and with Research Ethics Board approval to the principal investigators.

Cell Culture

Primary fibroblast cell lines were cultured in T25, T75, or T150 vented tissue culture flasks (Corning) with Dulbecco Modified Eagle Medium, containing 10% Fetal Bovine Serum, 100 µg/mL Pen/Strep, 50 µg/mL uridine and 100 µg/mL sodium pyruvate (Gibco) at 37 °C. Once fibroblast cells appeared confluent, cells were washed two to three times with sterile 1X Phosphate Buffered Saline (PBS) (Gibco). Cells were then trypsinized with two to four mL of 2.5% trypsin (Gibco), prewarmed to 37°C, and rocked

Table 2-1. Endometrial carcinoma samples.

Cancer Type	Sample Name	MSI	Grade	Age
Endometrial Carcinoma	RA26	Positive	1	77
Endometrial Carcinoma	RA8	Positive	1	58
Endometrial Carcinoma	AH161A	Positive	3	59
Endometrial Carcinoma	AH161B	Positive	3	59
Endometrial Carcinoma	AH153	Positive	3	84
Endometrial Carcinoma	AH225	Negative	2	72
Endometrial Carcinoma	AH223	Negative	1	62
Endometrial Carcinoma	AH221	Negative	1	75
Endometrial Carcinoma	RA2	Negative	3	64
Endometrial Carcinoma	RA3	Negative	3	57
Endometrial Carcinoma	RA4	Negative	3	84
Endometrial Carcinoma	TA	Positive	IIA	50
Matched Normal	NA	n/a	n/a	50
Endometrial Carcinoma	TB	Positive	IB	52
Matched Normal	NB	n/a	n/a	52
Endometrial Carcinoma	TC	Positive	IC	64
Matched Normal	NC	n/a	n/a	64
Endometrial Carcinoma	TD	Negative	IB	55
Matched Normal	ND	n/a	n/a	55
Endometrial Carcinoma	TE	Negative	IA	68
Matched Normal	NE	n/a	n/a	68

All endometrial carcinoma samples are of the endometriod subtype. The sample identification, microsatellite instability (MSI) status, age of patient and grade of tumour are given in the table. Microsatellite status and grade of tumour are not applicable to the matched normal samples and are designated as n/a. Two different scales are used to determine the grade of endometrial tumours. In the first scale, tumours are ranked from one to three, with a grade of one being the least severe and representing a tumour with less than 5% solid growth, a grade of two representing between 5 and 50% solid growth, and a grade of three, being most severe, representing greater than 50% solid growth. The second scale ranks tumours from IA to IIIC, and represents both the origin of tumour development as well as sites of metastasis. In the table above, grade IA represents a tumour that is limited to the endometrium, IB represents a tumour that has invaded into less than half of the myometrium, IC represents a tumour that has invaded into more than half of the myometrium, and IIA represents involvement of the endocervical glands and surface.

Table 2-2. Multiple myeloma cancer samples.

Cancer Type	Sample Name	Stage	Treatment	Age
Multiple Myeloma	GEH5	End	Chemotherapy	73
Multiple Myeloma	TTA2BM	Advanced	Untreated	47
Multiple Myeloma	WEB1BM	Advanced	Untreated	40
Multiple Myeloma	AJU15BM	Advanced	Chemotherapy	70
Multiple Myeloma	VIG1BM	Advanced	Untreated	48
Multiple Myeloma	DIK1BM	Intermediate	Untreated	60
Multiple Myeloma	BPA1BM	Early	Untreated	60
Multiple Myeloma	JCO1BM	Unknown	Untreated	54

The sample identification, stage, treatment and patient age are given in the table. Stages progress from early, being the least severe, through intermediate, advanced and end stage.

gently for one to three minutes. Cells were released from the flask by a sharp tapping motion, approximately eight mL of culture media was quickly added to halt the trypsin digestion, and the cells were redistributed to two or more new culture flasks.

Fibroblast cell lines lacking mtDNA (ρ^0) (Dr. Brian Robinson, Hospital for Sick Children, ON) were grown in a similar manner, but the culture media was supplemented with 20% Fetal Bovine Serum.

DNA Purification from Cultured Cells

Cells to be harvested were washed twice with sterile 1X PBS and trypsinized as described previously. Trypsin was neutralized with seven mL of culture medium. The cells were placed in a 50 mL tube (Sarstedt) and centrifuged at 3000 rpm for 10 minutes. The culture medium was removed using a pipette and the cell pellet was resuspended in five mL 1X PBS. The cells were subsequently centrifuged at 2500 rpm for five minutes. The PBS wash was repeated twice to ensure complete removal of trypsin and culture media. The cell pellet was resuspended gently in 1-3 mL of 4°C lysis buffer (100mM Tris-Cl, pH 8.0, 50 mM EDTA, 100 µg/mL proteinase K), transferred to a 15 mL conical tube and incubated overnight at 55°C. Following the incubation, an equal volume of 0.1 M NaCl was added and mixed by inversion. Five mL of phenol/chloroform/isoamyl alcohol (25:24:1) mixture was added and mixed by vortexing. The tube was spun in the J6-MI centrifuge (Beckman) for at 4°C for five minutes at 4000 rpm. The top aqueous layer of the tube was removed and placed in a new 15 mL tube, and the phenol/chloroform/isoamyl alcohol extraction was repeated. Following the second extraction, the aqueous layer was again removed and placed in a fresh 15 mL tube and three volumes of cold 100% ethanol were added and mixed gently by inversion. The

mixture was centrifuged at 4°C for 30 minutes at 4000 rpm to obtain the crude DNA pellet. Following centrifugation, the supernatant was removed and the pellet was resuspended in 300 µL of sterile water and transferred to a 1.5 mL microcentrifuge tube. 1.2 mL of cold 100% ethanol was added, and the tube was spun in the microcentrifuge at 14000 rpm for five minutes. The DNA pellet was subsequently washed with 70% ethanol and dried using the SpeedVac Plus (Savant) for approximately 10 minutes. 50 µL of sterile water was used to resuspend the DNA.

PCR Amplification of HVI and HVII Mitochondrial DNA Fragments

Database analysis (MitBase, http://bighost.area.ba.cnr.it) was performed in order to determine the frequency of mtDNA polymorphisms. PCR primers were designed to encompass as many highly polymorphic sites as possible for each D-loop HV region, while being sure to include the C-tract length polymorphism at position 303 within the primer region for HVII. Primers were first tested on a ρ^0 cell line to ensure that amplification of nuclear pseudogenes did not occur (Parfait *et al.*, 1998). PCR was carried out using 150 - 300 ng of template DNA in a 25 µL reaction volume (10 ng/µL primer, 1.5 mM $MgCl_2$, 0.1 mM dNTPs, 1 unit Taq Polymerase (Invitrogen)). The HVI PCR primers encompassed nucleotides 16049-16402, and consisted of forward primer: GTACCACCCAAGTATTGA and reverse primer: CGGAGGATGGTGGTCAA. The HVII PCR primers encompassed nucleotides 135-469 and consisted of forward primer: TGATTCCTGCCTCATCCTAT and reverse primer: GGGAGTGGGAGGGGAAA (Figure 2-1). Reactions were cycled 30 times for 30 sec at 92°C, 40 sec at 55°C and 50 sec at 72°C. Controls performed with AmpliTaq Gold polymerase (Applied Biosystems) were performed as described above, but with 30 ng of template and 1 U AmpliTaq Gold

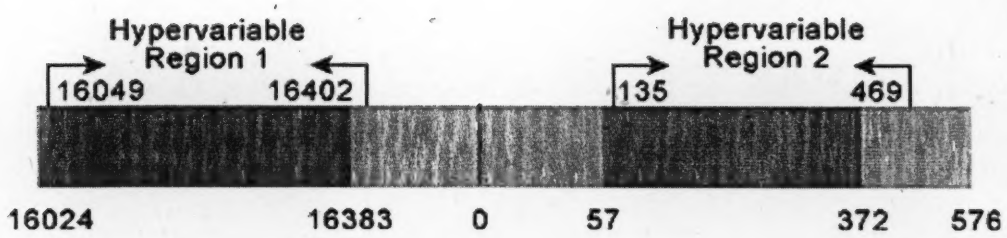

Figure 2-1. Schematic of the D-loop and the PCR primers within HVI and HVII. The D-loop encompasses nt 16042-576, and housed within the D-loop are two hypervariable regions, HVI and HVII, found at nt 16024-16383 and nt 57-372, respectively. The PCR primers utilized in this study were designed primarily to investigate the presence and frequency of polymorphism within the HV regions of the D-loop, and to investigate the relevance of the C303 marker found within HVII.

polymerase and a denaturing temperature of 95°C. Five µL of the amplified PCR product were visualized on a 1% (w/v) agarose gel with ethidium bromide under UV light. The remaining 20 µL of the amplified PCR product were purified as per the manufacturer's instructions using the QiaQuick PCR Purification Kit (Qiagen).

Ligation and Transformation of Mitochondrial DNA PCR Products

As conventional PCR-based automated sequencing approaches are reported to detect mtDNA heteroplasmy only above 30%, we adopted a subcloning-based sequencing approach to increase the limit of detection to ~10% (Tully *et al.*, 1999). One µL of purified HVI and HVII PCR product was ligated into the pGEM-T Easy Vector System overnight at 4°C (Promega) (Figure 2-2). Aliquots of competent DH5α *E. coli* cells were thawed on ice. Two µL of plasmid DNA were placed in a 1.5 mL microcentrifuge tube on ice, and 100 µL of the competent cell suspension were added to the plasmid DNA and the mixture was then incubated on ice for 30 minutes. The cells were then heat-shocked at 42°C for 45 seconds and placed on ice for two minutes. 900 µL of liquid LB medium was added to each of the cell mixtures, which were subsequently transferred to sterile glass tubes and incubated at 37°C for 90 minutes with shaking (250 rpm). The cells were plated on solid LB/Ampicillan/X-Gal/IPTG plates to allow for antibiotic selection and blue/white screening of transformants, and grown overnight at 37°C. White positive transformants were used to inoculate three mL of LB/Ampicillan liquid media and grown overnight with shaking (250 rpm) at 37°C in 15 mL plastic tubes (Simport).

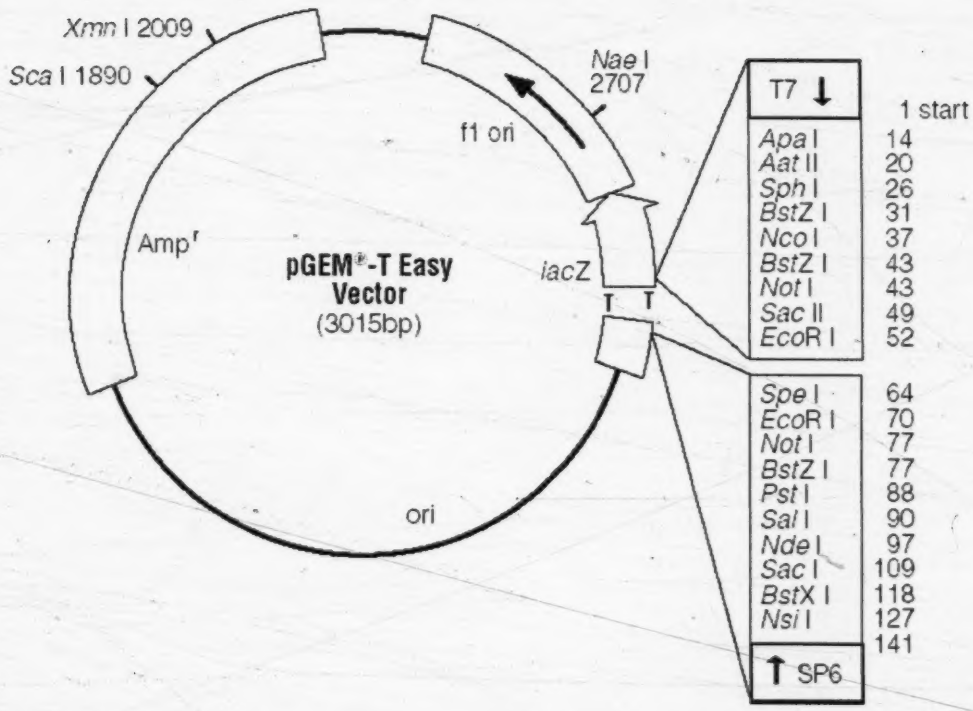

Figure 2-2. Schematic of the pGEM-T Easy PCR cloning vector. *Eco*RI sites within the multiple cloning site of pGEM-T Easy (Promega) allow removal of the PCR fragment of interest. Source of Figure: Promega, www.promega.com.

Plasmid DNA Purification

Plasmid DNA isolation was performed using the GenElute Plasmid Miniprep Kit (Sigma) and plasmid DNA was eluted in 120 µL of sterile water. If larger scale plasmid preparations were required, Triton Plasmid DNA Preparations were performed as described here. *E. coli* strains were plated onto solid LB media plates and incubated overnight at 37°C. Using a sterile spatula, the bacterial cells were scraped from the surface of the confluent plate and resuspended in one mL of sucrose buffer (5% sucrose, 50mM Tris, pH 8.0). One mL of lysozyme buffer (4.5 mg/mL lysozyme, 22.7 mM Tris pH 8.0, 113.6 mM EDTA, 0.9 mg/mL RNAse A) was added to the suspension, mixed, and placed on ice for 30 minutes. Following the incubation, one mL of cold Triton X-100 lysis buffer (0.3% Triton X-100, 0.185 M EDTA, 0.15 M Tris pH 8.0) was added, mixed by inversion, transferred to an ultracentrifuge tube and centrifuged in an Optima TLX ultracentrifuge (Beckman) at 40000 rpm at for 20 minutes at 4°C. The supernatant was then transferred to a 15 mL conical tube and an equal volume of water-saturated phenol was added. The contents were mixed by vortexing and centrifuged in a J6-MI centrifuge (Beckman) at 4200 rpm for five minutes at 4°C. The upper aqueous phase was removed with a Pasteur pipette and placed in a 15 mL tube. Ether washes were performed by filling the tube with ether and shaking the mixture vigorously. After the two phases had separated, the top layer of ether was removed with an aspirator. The ether wash was performed a minimum of three times. Once the lower phase was clear, 0.05 volumes of 5 M NaCl and 3 volumes of cold 100% ethanol were added to the lower phase and mixed by inversion. The tube was centrifuged at 4000 rpm for 10 minutes at 4°C, yielding a small oily pellet at the bottom of the tube. The supernatant was removed

from the pellet using an aspirator. The pellet was then dissolved in two mL of 2 M ammonium acetate, three volumes of cold 100% ethanol were added to the tube to precipitate the DNA, and the mixture was centrifuged at 4000 rpm for 10 minutes at 4°C. The white DNA pellet was dissolved in 0.3 mL of 2 M ammonium acetate and transferred to a 1.5 mL microcentrifuge tube. 2.5 mL of 100% ethanol was again added to re-precipitate the DNA and the tube was centrifuged at 14000 rpm. The supernatant was removed from the resulting DNA pellet by aspiration, and the pellet was washed first with 80% ethanol/0.2 mM EDTA, followed by a second wash with 80% ethanol. The DNA was dried in the SpeedVac Plus (Savant) for approximately 10 minutes and resuspended in 50 µL of sterile water. The concentration of the plasmid DNA was determined using the dna95 software on a Shimadzu UV-Visible UV-1601 PC spectrophotometer. *Eco*RI enzymatic digestion of seven µL of plasmid DNA was performed in order to confirm the presence of the mtDNA D-loop insert. Digest products were visualized on a 1% (w/v) agarose gel and visualized with ethidium bromide under UV light.

Li-Cor Sequencing

Ten or more positive transformants were selected for sequencing with M13 fluorescently labeled primers in order to determine the position and frequency of polymorphisms and the degree of heteroplasmy for each sample. M13 sequencing primers included the M13 Forward Primer: GTAAAACGACGGCCAGTG and the M13 Reverse Primer: GGAAACAGCTATGACCATG. Sequencing conditions included 30 seconds at 92°C to denature the template, 30 seconds at 60°C to allow M13 sequencing primer annealing, and a 30 second 72°C extension, repeated for 30 cycles. Samples were

either analyzed on a Li-Cor LongReadIR sequencing system (Li-Cor Biosciences) or sent to the DNA Sequencing Facility at the Centre for Applied Genomics (Toronto ON). Data were subject to statistical analysis using Fisher exact, chi-square and correlation tests (Moore *et al.*, 1993). Fisher exact tests were used to determine whether the mean values between two groups were significantly different, with neither group acting as a "control" group; whereas chi-square tests were used to ascertain whether a test group was statistically different from a control group. For Fisher exact tests and chi-square tests, a p-value greater than 0.05 was considered statistically significant. Correlation tests were used to determine the relationship between two distinct groups of samples (Moore *et al.*, 1993).

Microfluidic Chip Experiments

Autoclaved, filtered Milli-Q water was used to flush the microfluidic chip (Micralyne) for approximately two minutes before each chip load. The channels were loaded with approximately one μL of 5% Genescan polymer (Applied Biosystems) with 10% glycerol and 0.2 μM Sytox Orange DNA intercalating dye (Molecular Probes). The buffer reservoir and sample waste wells (CB1 and CB2, respectively) were loaded with three μL of 1X Tris Borate EDTA (TBE) buffer with 10% glycerol, and the buffer waste well (CB3) was loaded with three μL of 1X TBE buffer with 10% glycerol and 0.2 μM Sytox Orange. As the Sytox Orange dye is positively charged, when a voltage is applied, the dye travels up the channel, replenishing the intercalating dye that becomes bound to the separated DNA. The chip was analyzed for obstructions at the mouth of the well and at the channel intersection. 1.5 μL of *Eco*RI enzyme mixture (0.15 μL *Eco*RI enzyme, 0.15 μL 10X *Eco*RI buffer, 0.03 μL of RNAse A (New England Biolabs)) and 1.5 μL of

0.1 μg / μL plasmid DNA (pure or mixed in equal proportions) were loaded into the sample well (CB0) and mixed well by pipetting. Digestion was allowed to proceed at room temperature for 10 minutes, being sure to keep the chip covered to protect the Sytox Orange from photobleaching. Following digestion, 0.3 μL of 1X TBE with 10% glycerol was added to CB0, and CB1, CB2, and CB3 were partially replenished with 1 μL of their corresponding buffers. The chip was then loaded into the μTK, the apparatus that allows application of varying voltages, and hence separation and analysis of DNA, with the laser aligned 76 mm down the separation channel. A 60 second 0.3 kV injection with a positive voltage applied to sample waste well (CB2), followed by a 180 second 6.00 kV separation with a positive voltage applied to waste well (CB3), allowed detection of enzymatic digest products. DNA was detected in the separation channel at a wavelength of 532 nm, 76 mm from the intersection.

To further analyze restriction digest products by heteroduplex analysis, 1.5 μL of the reaction mixture was removed from the sample well, and replaced with two μL of formamide (Sigma). The addition of formamide allows the formation of single stranded DNA within the sample well (CB0). Formamide-treated mtDNA was injected for 60 seconds at 0.4 kV and separated for 180 seconds at 6.00 kV. Following injection, the formamide remains within the sample well while the mtDNA travels into the injection channel, facilitating re-annealing of the single stranded DNA and the formation of heteroduplexes at positions where sequence variation was present. For subsequent injections, the time was reduced to five seconds. Each formamide treatment was performed in triplicate to ensure reproducibility.

Chapter 3: Results

A portion of this chapter will contribute to the following publication:

Helmle, K.E., Dechant, A., Goodfellow, P.J., Backhouse, C.J., and Glerum, D.M. Analysis of mitochondrial DNA D-loop polymorphisms in human cancer. (Manuscript in preparation).

Analysis of D-loop Status in Normal Individuals

Normal Samples are Highly Polymorphic

Recently, several reports have examined the presence and frequency of mtDNA polymorphisms in individuals with different cancers (reviewed in Bianchi *et al.*, 2001; Carew *et al.*, 2002). The literature generally reports an increase in mtDNA mutation and polymorphism, especially within the D-loop regulatory region, and often the sequence variants are found at homoplasmic levels. The incidence of sequence variation in cancer samples has not been effectively compared to the frequencies of these polymorphisms observed in the normal population. In order to establish the significance of the presence and frequency of polymorphism within the D-loop control region, a subcloning-based sequencing approach was used to map sequence changes (relative to the Cambridge reference sequence (Anderson *et al.*, 1981)) in 353 bp and 334 bp segments of HVI and HVII, respectively. As the limit of detection for direct sequencing of mtDNA from a PCR product has been reported as 30% or greater, 10 or more subclones, chosen at random, from each PCR amplified sample were sequenced. This subcloning-based sequencing approach should increase sensitivity by as much as 20%, as sequence changes present in ~10% of the mtDNA population would be detected (Tully *et al.*, 1999).

A total of 475 normal HVI and HVII subclones were sequenced, derived from 12 blood controls obtained from normal individuals, five normal blood controls from patients with endometrial tumours and two fibroblast cell lines derived from normal individuals with normal mitochondrial function. Overall, the normal samples were very polymorphic, with a total of 67 different polymorphisms detected within the 687 bp of D-loop sequence studied (Figure 3-1). Between two and 20 polymorphisms were recorded

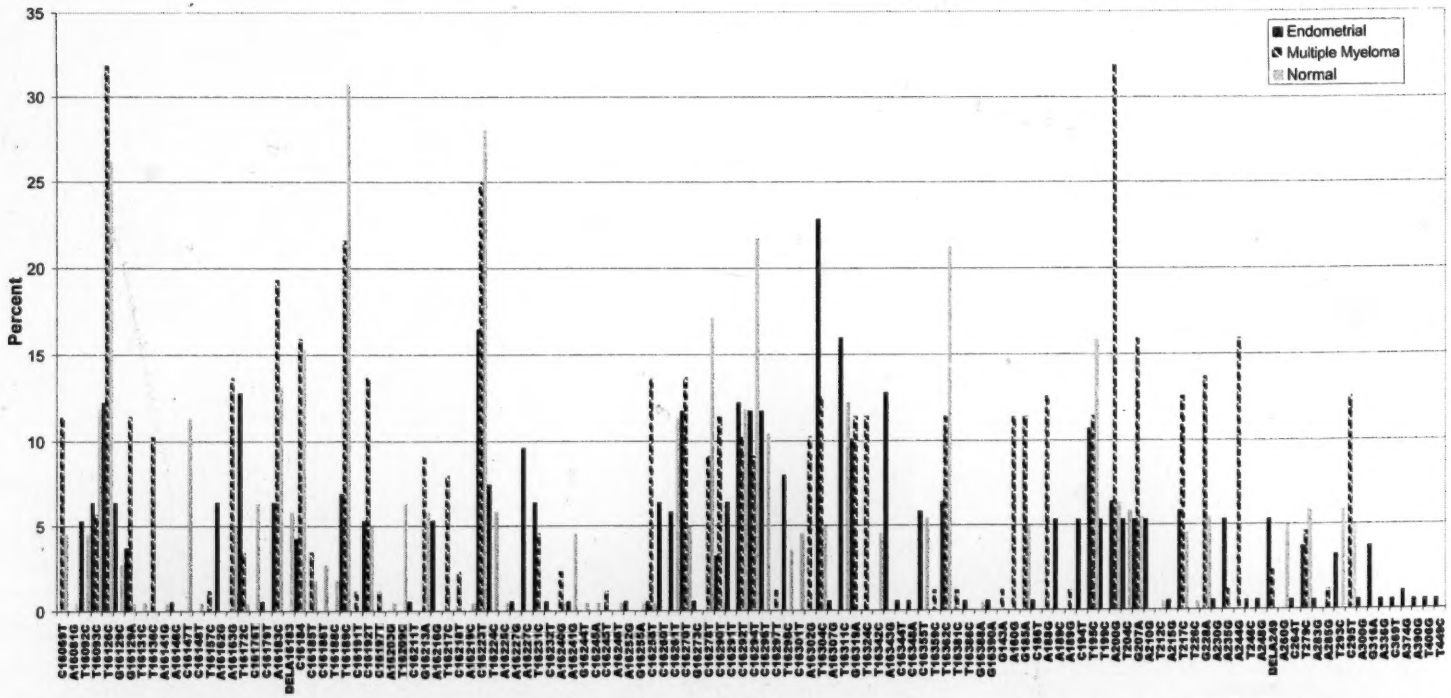

Figure 3-1. Frequency of mtDNA polymorphism in endometrial, multiple myeloma and normal subclones. The site and frequency of polymorphism are given in the figure, with endometrial, multiple myeloma and normal samples represented by solid black, hatched, and solid grey bars, respectively. The height of the bars reflects the frequency of polymorphism in a given group of samples, and was determined by dividing the total number of polymorphic subclones at a given site by the total number of subclones for each sample type. A263G and insC311 are omitted as they appear in 100% of the subclones analyzed, regardless of sample type.

across the HV regions in any given sample. Samples had an average of 7.3 and 3.2 polymorphisms for HVI and HVII, respectively, giving an average frequency of 1 polymorphism every 50 base pairs in HVI and every 100 base pairs in HVII. While most polymorphisms have been published or reported in mtDNA databases (Mitomap and MitBase), seven previously unreported sequence changes were detected in normal samples (Table 3-1). The majority of unreported polymorphisms described here are transitions, consistent with the propensity for transitions observed in the mitochondrial genome, which are caused primarily by the deamination of cytosine residues.

In order to ensure that sequencing 10 subclones would be representative of both the presence and frequency of the polymorphisms present in a given sample, a total of 51 subclones were sequenced for one of the control samples to determine whether 10 subclones provided a sufficient sample size. Upon sequencing, no significant difference was observed when compared to the results obtained by sequencing only 10 subclones (Figure 3-2). Two additional polymorphisms, T204C and T226C, both reported, were observed in single, distinct subclones. The C-tract length at position 303 varied slightly when 51 subclones were sequenced, with the C-tract spectrum ranging from seven to nine cytosines when 10 subclones were sequenced, and ranging from seven to 10 cytosines when more than 50 subclones were sequenced. The relative proportions of the C-tract length components remained similar in the larger group of subclones when compared to the group of 10. I therefore concluded that sequencing 10 subclones was representative and would be sufficient to approximate the frequency of polymorphism.

Furthermore, to address the possibility that a PCR error by the Taq DNA polymerase would be misinterpreted as a polymorphism, several sequences were re-

Table 3-1: Unreported polymorphisms.

Polymorphism	Sample
A160G	Multiple Myeloma
A244G	Multiple Myeloma
A260G	Normal
A283G	Endometrial
T293C	Normal, Endometrial
A300G	Endometrial
A336G	Endometrial
C369T	Endometrial
T406G	Endometrial
T449C	Endometrial
A16081G	Normal
A16141G	Normal
A16146C	Endometrial
A16225C	Normal
A16227C	Endometrial
G16273C	Endometrial
A16307G	Endometrial
G16244T	Normal
C16245A	Normal
C16348A	Endometrial
T16386C	Endometrial

Upon sequencing of 475, 373 and 179 subclones derived from normal, endometrial and multiple myeloma samples, respectively, a total of 125 different polymorphisms were detected, 21 of which were previously unreported in the Mitomap or Mitbase mtDNA databases and are listed above.

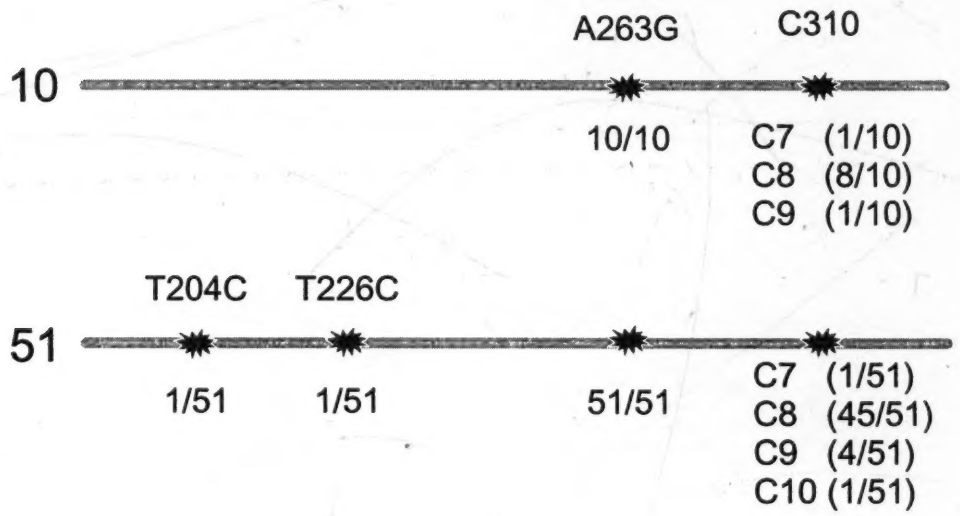

Figure 3-2. Schematic of sequencing results for ten subclones compared to 50 subclones of a normal blood control. The polymorphism and its frequency are indicated above and below the mtDNA schematic, respectively.

amplified with AmpliTaq Gold (Table 3-2). In all cases, the distribution of polymorphisms using AmpliTaq Gold was similar to the initial sequences obtained using Taq DNA polymerase; no new polymorphisms were detected, and all expected polymorphisms appeared at a frequency comparable to that of the original sequence. In several instances, polymorphisms at a frequency of 1/10 were not reproduced upon sequencing with AmpliTaq Gold; however, only 10 additional subclones were sequenced with the more sensitive Taq polymerase, and it is statistically reasonable that polymorphisms representing rare populations would not be isolated a second time. Based on this observation, I assumed that the likelihood of PCR-based errors in this study was very low.

In utilizing a subcloning based sequencing approach, the presence of 'clonal populations' was easily discernable within samples. Often polymorphisms spanning a region of over 250 bp appeared linked, and different populations of mtDNA molecules could be identified. For most samples, two or three distinct populations of mtDNA molecules were detected, although up to nine different populations were present in some cases. Although these linked polymorphisms could be the result of established mtDNA haplotypes, the polymorphisms do not follow an apparent trend, and are unique to the sample. In several cases, an unreported polymorphism appears linked to a reported polymorphism, and is detected in more than one subclone of a given sample.

Normal Samples are Primarily Heteroplasmic

Most reports in the literature indicate that normal individuals tend to be homoplasmic, and that heteroplasmy is characteristic of a mitochondrial DNA disorder (Monnat *et al.*, 1985). Recent literature, however, indicates that heteroplasmy may be

Table 3-2. Samples re-amplified with AmpliTaq Gold.

Fibroblast sample R	AmpliTaq R	Normal blood 2694	AmpliTaq 2694	Normal blood 2697	AmpliTaq 2697	Endometrial tumour AH221	AmpliTaq AH221
T16093C (11/11)	T16093C (10/10)	*A16141G (1/13)*		A263G (10/10)	A263G (10/10)	A189C (10/10)	A189C (10/10)
A16183C (1/11)	A16183C (2/10)			C7 (1/10) C8 (8/10) C9 (1/10)	C8 (8/10) C9 (1/10) C10 (1/10)	C194T (10/10)	C194T (10/10)
delA16183 (8/11)	delA16183 (4/10)			insC311 (10/10)	insC311 (10/10)	T195C (10/10)	T195C (10/10)
insC16184 (4/11)	insC16184 (3/10)					T199C (10/10)	T199C (10/10)
insCC16184 (1/11)	insCC16184 (3/10)					T204C (10/10)	T204C (10/10)
T16189C (11/11)	T16189C (10/10)					G207A (10/10)	G207A (10/10)
G16213A (11/11)	G16213A (10/10)					A210G (10/10)	A210G (10/10)
C16223T (11/11)	C16223T (10/10)					*A230G (1/10)*	
A16225C (1/11)						A263G (10/10)	A263G (10/10)
C16278T (11/11)	C16278T (10/10)					*A283G (1/10)*	
						C8 (9/10) C10 (1/10)	C8 (8/10) C9 (1/10) C10 (1/10)
						insC311 (10/10)	insC311 (10/10)
						C369T (1/10)	
						A374G (1/10)	

Original sequence results and the results obtained upon resequencing with high fidelity AmpliTaq Gold are reported in the table. C7, C8, C9 and C10 represent C303 lengths. The frequency of each polymorphism is given in parentheses, and italicized values represent polymorphisms that are not reported in mtDNA databases.

detected in normal individuals, but its reported incidence is rare (Comas *et al.*, 1995; Kirches *et al.*, 2001). In our cross-section of normal controls, varying from highly glycolytic fibroblast cells to normal blood samples, it was determined that heteroplasmy is common. Across HVI and HVII, 63% and 47.4% of the control samples, respectively, were heteroplasmic (Table 3-3). The number of normal samples that are homoplasmic across both HV regions examined was 3/19 (15.8%). It is important to note that the primer sets utilized in this study were designed to encompass as many frequently polymorphic sites as possible, with the idea that these sites would be the most likely to display heteroplasmy in a sample if it were present. The definition of "homoplasmy" implies that the sequence of the entire mtDNA molecule is identical; therefore, the levels of heteroplasmy within a sample are likely underestimates, as only a small fraction of highly polymorphic sites are being examined and the remaining ~16kb of mtDNA has not been investigated.

The higher level of heteroplasmy I have identified is likely due to the subcloning-based sequencing approach. Direct sequencing of PCR products identifies heteroplasmy only if the frequency of the mtDNA polymorphism lies between 30% and 70%. Frequency levels below 30% may falsely indicate a homoplasmic wildtype sequence, whereas frequencies greater than 70% potentially indicate a homoplasmic polymorphism. In order to compare my results with those from conventional PCR-based sequencing methods, the number of polymorphisms detected here that fell within the 30-70% window was calculated, indicating that only 10.4% of the polymorphisms detected in normal samples would have been detected using direct sequencing techniques.

Table 3-3. Incidence of homoplasmy in normal, endometrial, and multiple myeloma samples

Sample	HVI	HVII	HVI and HVII	Sample Size
Normal	36.8%	52.6%	15.8%	19
Endometrial	31.3%	37.5%	18.8%	16
Multiple Myeloma	25.0%	12.5%	12.5%	8

The incidence of homoplasmy within HVI or HVII varies across the groups of samples studied; however, the number of samples exhibiting homoplasmy across both HV regions is consistently low, independent of cancer status. Chi-square analysis demonstrates that there is no association between the presence of cancer and homoplasmy within the tumour mtDNA when compared to normal controls ($p > 0.90$).

Analysis of D-loop Status in Cancer Samples

The Level of Polymorphism in Cancer Samples is Similar to Normal Controls

In order to determine the level of polymorphism in cancer samples, the hypervariable regions of 16 endometrial tumours (183 HVI and 188 HVII subclones) and 8 multiple myeloma samples (91 HVI and 88 HVII subclones) were sequenced. 76 different polymorphisms were detected in the endometrial carcinomas, 13 of which were previously unreported in either the Mitomap or MitBase databases (Figure 3-1, Table 3-1). An average of 7.5 polymorphisms per sample was detected across both hypervariable regions. Multiple myeloma samples displayed a total of 55 polymorphisms, only two of which were previously unreported (Figure 3-1, Table 3-1). Multiple myeloma samples showed an increased frequency of polymorphism, with an average of 12.6 polymorphisms per sample; however, statistical analysis demonstrated that the number of polymorphisms present in either endometrial cancer or multiple myeloma was not significantly higher than the frequency of polymorphism for the normal control group ($p = 0.468$ for endometrial carcinoma, and $p = 0.137$ for multiple myeloma, Fisher exact test). Furthermore, there was no statistical correlation between the frequency of polymorphism and grade of cancer in ($p > 0.1$, Fisher exact test). As all endometrial cancer patients and all but two multiple myeloma patients were untreated at the time the tissue samples were obtained, additional data would be required for statistical analysis of the effect of cancer treatment on the frequency of mtDNA polymorphism.

Two reported polymorphisms, A263G and the insertion of a cytosine at position 311, were present in all 487 HVII subclones analyzed, regardless of sample status. The prevalence of these polymorphisms likely indicates an error in the reported reference

sequence at sites 263 and 311. Interestingly, the Anderson reference sequence originally cited a guanine at position 263, although the databases have since changed this to an adenine, stating that the A-G polymorphism is rare. The presence of six cytosines rather than five at position 311 has been reported in the literature previously (Kirches *et al.*, 2001).

Many of the unreported polymorphisms were found in a single subclone of a single sample and likely reflect the detection of rare polymorphisms, representing a very small percentage of the total population of mtDNA in the sample. However, some unreported polymorphisms were found in more than one subclone, and often unreported polymorphisms were detected in more than one sample or even between groups of samples. T293C, an unreported polymorphism that was detected in 6/14 subclones from one endometrial sample, was present in only 1/12 subclones in a normal control, but was homoplasmic in a second normal control. A244G was found in two multiple myeloma samples, where it was homoplasmic in one sample and found in 2/12 subclones of the second sample. Furthermore, several unreported polymorphisms appeared in more than one subclone of a given sample, and often appeared linked to other polymorphisms, both reported and unreported. In an extreme case, one endometrial sample displayed five unreported polymorphisms; three unreported polymorphisms were detected within a single subclone, and two other unreported polymorphisms, linked to a reported polymorphism, were found on a second subclone. It is unlikely that Taq errors would occur in such an organized fashion, suggesting that the unreported polymorphisms detected here legitimately represent the mtDNA sequence of the samples in question.

The position and frequency of polymorphism varies considerably within samples and within groups of samples (Figure 3-1). With the exception of the A263G and insC311 polymorphisms, no single polymorphism has a frequency higher than 35% in any of the groups studied. Furthermore, neither the endometrial cancer nor the multiple myeloma samples seem to exhibit an increase in the frequency of polymorphism across the two hypervariable regions when compared to the normal controls. The A260G polymorphism appears in a greater number of multiple myeloma subclones than endometrial or normal subclones, and a number of polymorphisms downstream of nt 260 within HVII appear at greater levels in multiple myeloma samples as well. Furthermore, certain low frequency polymorphisms at or downstream of nt 300 appear exclusively in endometrial samples. In many instances, however, the frequency of polymorphism is much greater in the normal samples than in the cancer groups.

To address variation between tumour samples and normal tissue from a single individual, ten or more subclones were sequenced from five endometrial tumours and their corresponding normal blood. Sequence results from HVI revealed a different tumour-specific variant in four different sample pairs (Table 3-4). In sample pair NA/TA (where samples beginning with N represent normal blood samples and samples beginning with T represent endometrial cancer samples), however, four variants were observed in the normal blood, but not in the corresponding tumour. All polymorphisms, both normal-specific and tumour-specific, occurred in a single subclone in each sample. There was no apparent trend that allowed us to discriminate between normal and tumour sample types. Upon examination of HVII, three of the five sample pairs were identical at all sites,

Table 3-4. HVI sequence results for matched normal and endometrial cancer samples

	T16092C	T16126C	G16129C	A16163G	A16183C	insC16184	delC16184	T16189C	C16192T	A16219C	C16232T	A16241G	C16245A	A16246T	C16256T	C16261T	C16270T	G16273C	C16292T	C16294T	C16296T	T16304C	T16311C	C16355T	T16362C
NA				*1/11*					11/11	*1/11*			*1/11*	*1/11*	11/11		11/11								
TA									10/10						10/10		10/10								
NB		12/12																	12/12	12/12	12/12			12/12	
TB		11/11										*1/11*							11/11	11/11	11/11			11/11	
NC		11/11														11/11				11/11	11/11	11/11			
TC		11/11									*1/11*					11/11				11/11	11/11	11/11			
ND																									
TD																		*1/12*							
NE	10/10		6/10		10/10	2/10	3/10	10/10																	10/10
TE	12/12		12/12		12/12	7/12	1/12	12/12															*1/12*		12/12

The site of polymorphism is listed according to the mtDNA reference sequence. Sample names beginning with N indicate a normal blood sample, and sample names beginning with T indicate an endometrial tumour sample. The frequency of polymorphism is given as the number of subclones containing a given polymorphism over the total number of subclones for a sample. Italicized frequencies represent polymorphisms that are normal- or tumour-specific.

Table 3-5. HVII sequence results for matched normal and endometrial cancer samples

	T195C	A200G	T212C	T217C	A263G	C204T	T293C	C303	insC311	G316A	A336G	A390G	T406G	T449C
NA					10/10			C8 10/10	10/10					
TA					11/11	*1/11*		C8 9/11 *C9 2/11*	11/11	*7/11*				
NB					11/11			C8 10/11 C9 1/11	11/11					
TB					10/10			*C7 2/10* C8 6/10 C9 2/10	10/10					
NC					11/11			C8 11/11	11/11					
TC					11/11			C8 11/11	11/11					
ND	*1/12*	1/12	*1/12*		12/12		*1/12*	C7 3/12 C8 7/12 *C9 2/12*	12/12					
TD		1/12			12/12			C7 8/12 C8 4/12	12/12		*1/12*	*1/12*	*1/12*	*1/12*
NE				10/10	10/10			C7 11/11	10/10					
TE				11/11	11/11			C7 11/11	11/11					

The site of polymorphism is listed according to the mtDNA reference sequence. The frequency of polymorphism is given as the number of subclones containing a given polymorphism over the total number of subclones for a sample. Italicized frequencies represent polymorphisms that are normal- or tumour-specific.

including the C-tract length polymorphism at position 303 (Table 3-5). I identified a polymorphism at nt 316 in one endometrial tumour sample (TA) in which a guanine was converted to an adenine in 7/11 subclones. This polymorphism was only observed in this single endometrial tumour, and was not present in any of the remaining normal or cancer subclones. I observed four normal-specific and four tumour-specific variants in the ND/TD pair, all at a frequency of 1/12, with the exception of the C9 variant in the normal ND sample which was present in 2/12 subclones. Overall, the cancer and normal samples showed an approximately equal incidence of polymorphism, with no observed trend distinguishing the two subsets of samples.

Reports in the literature have suggested an association between aging and the accumulation of large deletions or duplications within the D-loop region (Lee *et al.*, 1994). No large deletions (greater than two base pairs) or duplications were observed in the samples examined. There was no correlation between patient age and number of polymorphisms within the D-loop region for endometrial cancer patients ($r = 0.005$, Correlation Test). However, for multiple myeloma patients, a negative correlation was observed, indicating that younger patients tend to harbour more mtDNA polymorphisms than older patients ($r = -0.453$). Due to the small size of the multiple myeloma data set, the significance of this result requires further investigation.

The correlation between nuclear microsatellite instability (MSI) and the frequency of mtDNA polymorphisms was also investigated. The presence of nuclear MSI had been investigated previously for the endometrial carcinomas, with the presence of DNA MSI suggesting loss of mismatch repair within the nucleus (Dr. Susan Andrew, Department of Medical Genetics, University of Alberta; Dr. Paul Goodfellow, Department of Surgery,

Washington University School of Medicine, personal communication). Samples exhibiting nuclear MSI displayed an average of 4.6 polymorphisms per sample, whereas samples lacking nuclear MSI averaged 10.4 polymorphisms, over twice as many as the group lacking nuclear MSI ($p = 0.0122$, Fisher exact test).

The Degree of Homoplasmy is Comparable in Cancer Samples and Normal Controls

The propensity for homoplasmy with regard to mtDNA variants been reported as a characteristic of certain tumours and has been proposed as a marker for cancer diagnosis. When examining HVI alone, normal samples displayed the highest degree of homoplasmy, followed by endometrial and multiple myeloma samples (Table 3-3). Upon examination of HVII, again normal samples and multiple myeloma samples displayed the highest and lowest incidence of homoplasmy, respectively; however, the level of homoplasmy in normal samples was greatly increased, and decreased dramatically in multiple myeloma samples. Homoplasmy was found to a small degree in endometrial, multiple myeloma and normal samples overall, with values ranging from 12.5% in multiple myeloma to 18.8% in endometrial samples. When the two HV regions were taken together; approximately 85% of all samples tested were heteroplasmic. Moreover, there was no association ($p > 0.90$, Chi-square Test) between homoplasmy and the presence of cancer, and therefore statistically, normal and cancer samples are equally likely to appear homoplasmic.

When nuclear MSI status was taken into account, 3/8 samples exhibiting MSI and 2/8 samples lacking MSI were homoplasmic and therefore, no correlation was observed between nuclear MSI and the presence of homoplasmy. Furthermore, for the endometrial samples studied, there was no relationship between the presence of homoplasmy and

patient age (p = 0.129, Fisher exact test); the multiple myeloma patient sample size was too small to perform similar statistical analysis.

The C303 Polymorphism Varies Considerably in Normal Controls

The expansion or contraction of the cytosine tract at position 303 (C303 tract) has been suggested as a diagnostic and/or prognostic marker for carcinogenesis (Fliss *et al.*, 2000; Ha *et al.*, 2002; Parrella *et al.*, 2003; Sanchez-Cespedes *et al.*, 2001). Although it is assumed that the majority of normal samples have a tract of seven cytosines (C7) as presented in the mtDNA reference sequence, and that expansion above this level is rare, I determined during the course of this work that C-tract expansion is very common in normal controls, with the number of subclones exhibiting a C7 tract being approximately equal to the number exhibiting a tract of eight cytosines (C8) (Figure 3-3). In comparison, the majority of endometrial and multiple myeloma samples displayed a C8 tract. Each class of sample examined showed expansions up to C10. Overall, the majority of endometrial, multiple myeloma, and normal samples had an expanded C303 tract. The incidence of a C7 tract in normal subclones was significantly higher compared to endometrial subclones, but not compared to multiple myeloma samples (p = 0.01 for endometrial, p = 0.14 for multiple myeloma, Chi-square Test), whereas the incidence of a C8 tract was significantly lower in endometrial subclones, but not in multiple myeloma, when compared to normal controls (p = 0.01 for endometrial, p = 0.80 for multiple myeloma). The level of C303 tract expansion to C9 was significant in multiple myeloma (p = 0.001), but not in endometrial cancers (p = 0.1944). Finally, there was no significant difference in the C10 expansion in cancer samples when compared to normal controls (p

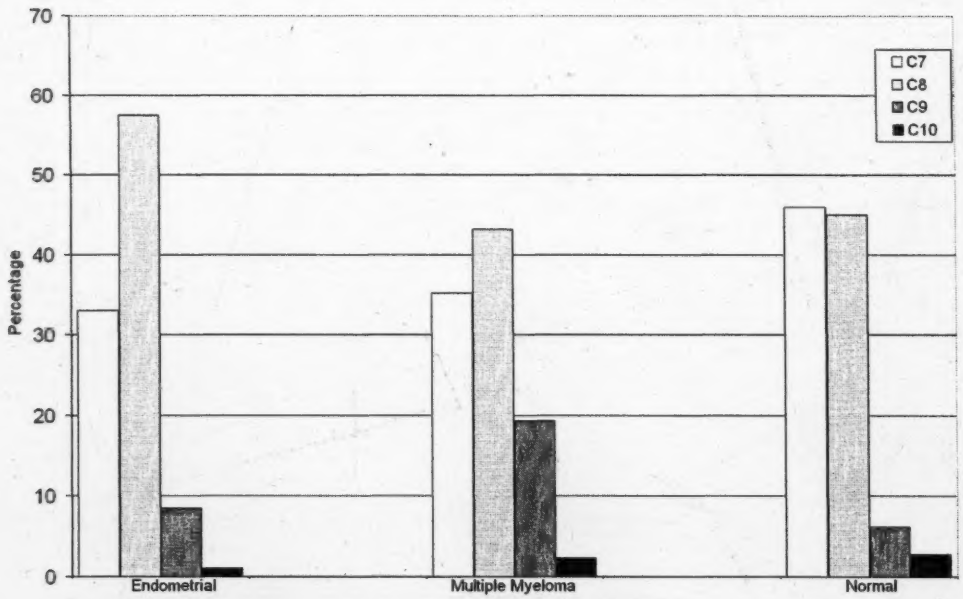

Figure 3-3. C303-tract length in normal, endometrial, and multiple myeloma subclones. The graph represents the proportion of subclones containing a C303-tract length of C7, C8, C9 and C10 compared to the total number of subclones for each sample class. White, light grey, dark grey and black bars represent C7, C8, C9 and C10, respectively.

= 0.15 for endometrial, p = 0.75 for multiple myeloma). Furthermore, Poisson analysis revealed that the distribution of C-tract lengths in cancer samples is comparable to the distribution of C-tract lengths in normal samples, with cancers showing only a slight increase in C9 length compared to the normal controls. In order to determine whether expansion above the 'normal' C7 tract is associated with cancer, the prevalence of the expanded sequences (C8-10) was compared in normal and cancer samples and found not to be significantly different between cancer samples and normal controls ($p > 0.05$).

Moreover, C303 tract expansions were often detected even within a single sample, which sometimes displayed as many as four different C303 tract lengths. Heteroplasmy at this site could be attributed to the heterogeneity of the sample; however, even microdissected endometrial cancer samples showed multiple populations at the C303 tract locus. This heteroplasmy was observed in both normal and cancer samples and likely reflects the normal state in mitochondria.

Development of Microfluidic Chip-Based mtDNA Assays

Restriction Digestion and On-Chip Labelling

In order to facilitate later mtDNA mutation detection using a microfluidic chip platform, a method of DNA self-labelling within the separation channel was first required in order to minimize any interference by DNA-binding dyes within the sample well. Sytox Orange (Molecular Probes, OR USA) is an intercalating dye that fluoresces when bound to double-stranded nucleic acid. The use of 0.2 μM Sytox Orange both in the waste buffer well and in the polymer allowed for successful integration of Sytox Orange on-chip DNA labelling and DNA electrophoresis. *Eco*RI digestion of the pGEM-T Easy plasmid vector containing the mtDNA D-loop HVI or HVII PCR product excises the

fragment of interest from the empty vector (Figure 2-2). *Eco*RI digestion was achieved using enzymatic digestion reagents at standard concentrations in a total volume of three μL. Increasing the amount of enzyme did not increase the efficiency of the restriction digest and lowering the total concentration of plasmid DNA to be digested resulted in a more efficient digest with stronger insert and empty plasmid peaks. The digestion progressed to a sufficient degree after a 10 minute, room temperature incubation as confirmed by analysis prior to digest.

The separation of digested mtDNA-containing plasmid vector DNA resulted in three distinct and reproducible peaks (Figure 3-4). As the undigested plasmid, retaining the mtDNA PCR insert is circular and supercoiled it electrophoreses through the sieving matrix at a much more rapid rate than is expected for a 3.3 kb molecule, and appears as a broad hump at approximately 140 seconds. The liberated mtDNA PCR insert (353 bp for HVI and 334 bp for HVII) appeared as a much sharper peak as its content is more uniform, and arrives at approximately 180 seconds. Finally, the 3.0 kb linearized plasmid appears as a distinct, sharp peak at approximately 205 seconds. Due to its length, and hence its increased ability to harbour Sytox Orange molecules, the linearized plasmid fluoresces most intensely of the three peaks. The intensity of the undigested plasmid peak and the insert peak vary with the efficiency of the digest. "Extra" peaks often appear at random during a run, but their peak width is typically smaller than that of a true peak. To ensure that an extra peak is not misinterpreted as an authentic peak, all runs were performed in triplicate. The cause of the appearance of extra peaks remains unclear.

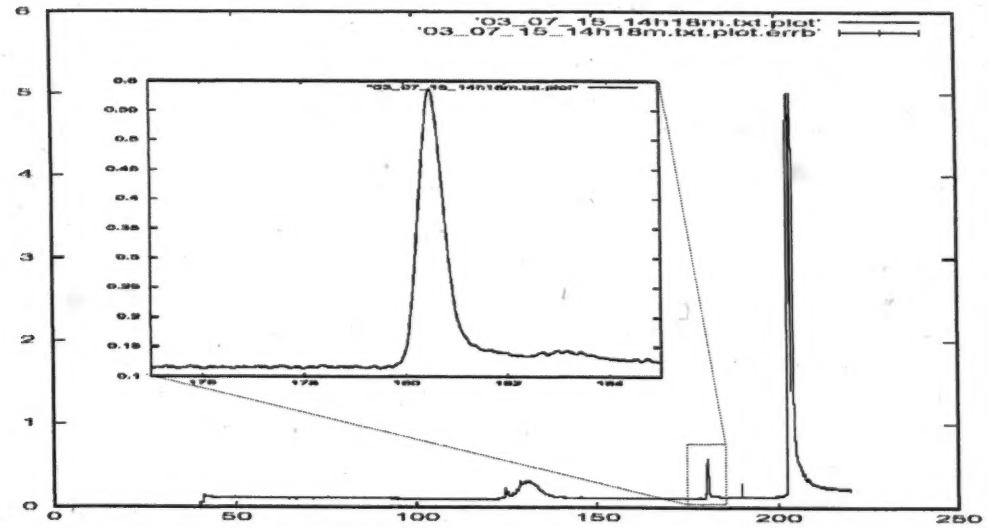

Figure 3-4. On-chip restriction digest with sytox orange labeling. Here, the undigested plasmid appears as a broad hump at approximately 130 seconds, the insert appears as a sharper peak at approximately 180 seconds, and the linearized plasmid appears as a sharp peak at approximately 205 seconds.

Heteroduplex Analysis

Heteroduplex analysis is a form of mutation detection that takes advantage of the intrinsic self-complementary nature of DNA to identify sequence variation. DNA is first denatured by heat or chemical agents, and then allowed to reanneal; if a mutation is present in the DNA population being examined, a bulge results when wildtype and mutant copies of single-stranded DNA attempt to reanneal. In order to perform successful microfluidic chip-based heteroduplex mtDNA analysis, excess formamide was added to the sample well following restriction enzyme digestion to allow temporary formation of single stranded DNA within the sample well. The DNA could then be injected into the channel, leaving the formamide in the sample well. DNA heteroduplexes should result within the injection channel when the digestion was performed on a 50:50 mixture of two mtDNA populations, and the heteroduplexes were electrophoresed through the separation channel and detected at 76 mm.

In order to develop the heteroduplex analysis protocol on-chip, a 50:50 mixture of two mtDNA subclones differing at ten distinct sites was subjected to heteroduplex analysis using excess formamide (Figure 3-5). Subclone 6-2 was derived from a normal fibroblast cell line and had three polymorphisms (T16126C, C16292T, and C16294T). Subclone 7-2 was derived from a second normal fibroblast cell line and displayed 7 polymorphisms, including the deletion of an adenine and the insertion of an adjacent cytosine, leaving the net length of the mtDNA insert unchanged (T16093C, delA16183, insC16184, T16189C, G16213A, C16223T, and C16278T). Heteroduplex analysis of this sample pair reveals a broad peak at approximately 90 seconds, representing undigested plasmid DNA, and the linearized vector peak at approximately 165 seconds.

As the vector lacks sequence variation amongst its molecules, and would ultimately contain very little common sequence with the mtDNA insert, the formamide treatment left the 3.0 kb sequences unaffected. Most importantly, the single sample peak observed after the restriction enzyme digestion (Figure 3-4) is transformed into a triplet peak following the formamide treatment (Figure 3-5). It is likely that the two homoduplexes representing the reformation of samples 6-2 and 7-2 migrate together, and the two heteroduplexes formed through the mispairing of 6-2 with 7-2 migrate separately and more slowly than the homoduplexes. The formation of heteroduplexes and their relative ease of identification provide "proof of principle" results indicating that heteroduplex analysis of mtDNA on-chip can ultimately be used in mutation detection.

Preliminary results with different combinations of subclones are limited; however, heteroduplex analysis has been performed on two additional mixtures. HVI subclones 6-2 (see sequence information above) and 7-1 (T16093C, delA16183, T16189C, G16213A, C16223T, and C16278T), both derived from normal fibroblast cell lines, differ at a total of nine sites. The polymorphism profile of 7-1 is nearly identical to 7-2, with the exception that the insertion of a cytosine does not occur within subclone 7-1, leaving a length polymorphism that is not present in the 6-2/7-2 mixture studied above (Figure 3-6). Interestingly, when the heteroduplex profile of 6-2/7-1 is obtained, the triplet structure of the 6-2/7-2 mixture is lost, and is replaced by a structure consisting of only two peaks. It appears that the mutation or polymorphism spectrum of the individual mixtures personalizes the resulting heteroduplex profile.

Finally, preliminary results from heteroduplex analysis of HVII subclones was obtained by mixing subclones 4-9 (A263G, T279C, C7, insC311) and 5-8 (T195C,

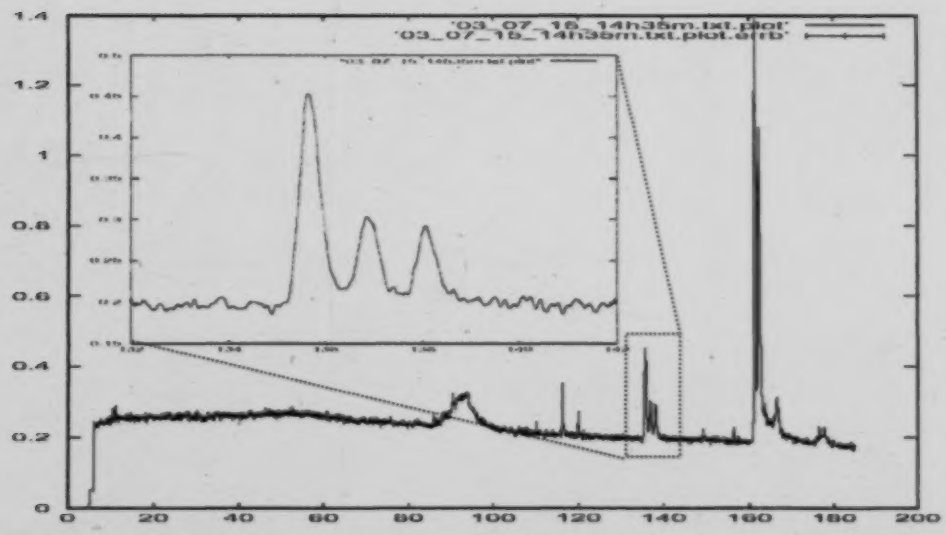

Figure 3-5. Heteroduplex analysis of a 50:50 population of mtDNA restriction digest products. The appearance of the undigested plasmid peak occurs at approximately 90 seconds, the heteroduplex triplet representing the mtDNA inserts appears at approximately 135 seconds, and the linearized vector peak arrives at approximately 165 seconds.

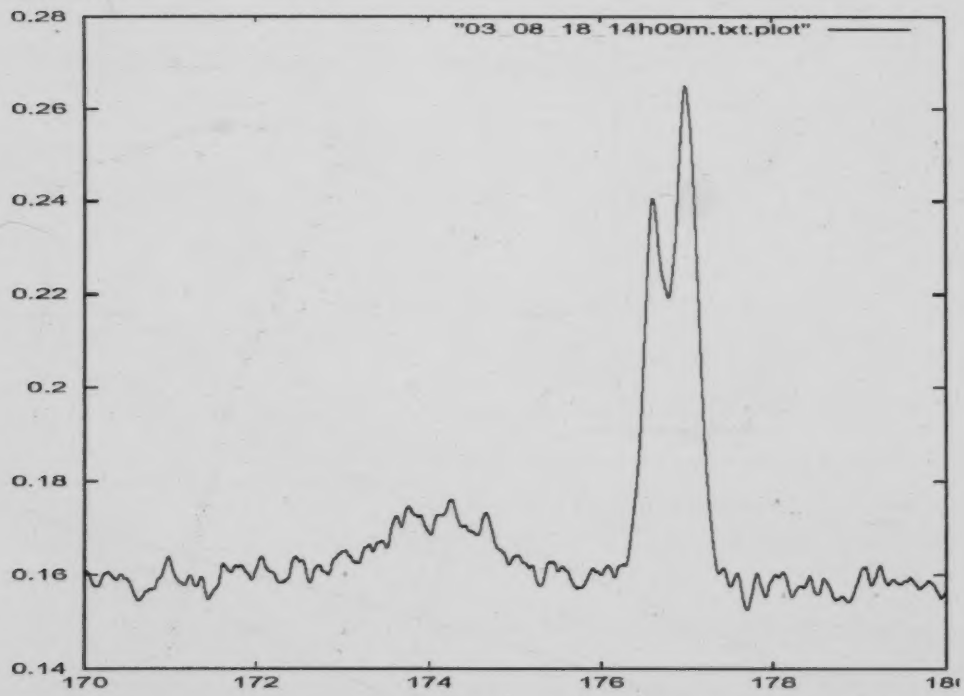

Figure 3-6. Heteroduplex analysis of HVI subclones 6-2 and 7-1. Subclones 6-2 and 7-1 were mixed in equal proportions and subjected to heteroduplex analysis. The duplex insert peak appears at 176 seconds. The uncut plasmid and linearized vector peaks are not shown.

A200G, A263G, C9, insC311) in equal proportions. The two subclones differ at a total of 4 sites, with subclone 5-8 having two additional cytosines when compared to sample 4-9. The additional cytosines represent a C303 length polymorphism. Consistent with previous results, the heteroduplex results appear unique to the combination of polymorphisms present, with the appearance of a triplet peak and the first two peaks resembling a single clefted peak (Figure 3-7).

The low signal intensity and the high, dropping baseline displayed in Figures 3-6 and 3-7, respectively, represent inconsistencies in the heteroduplex analysis protocol that require further troubleshooting. The baseline variations likely indicate an interaction between the Genescan polymer and the use of Sytox Orange intercalating dye. Furthermore, several additional combinations need to be analyzed before conclusive results regarding the nature of the heteroduplexes may be obtained.

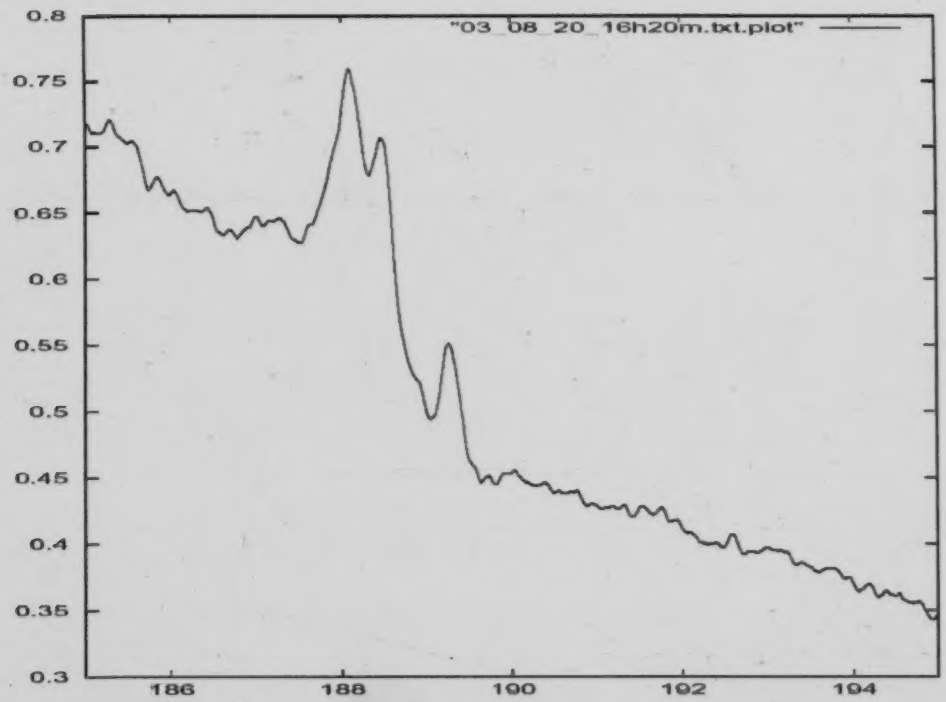

Figure 3-7. Heteroduplex analysis of HVII subclones 4-9 and 5-8. Subclones 4-9 and 5-8 were mixed in equal proportions and subjected to heteroduplex analysis. The clefted insert peak appears at 188 seconds. The uncut plasmid and linearized vector peaks are not shown.

Chapter 4: Discussion

Mitochondrial DNA Variation in Normal Samples

An increasing number of reports examining the role of mtDNA in cancer have been published recently (Bianchi *et al.*, 2001; Penta *et al.*, 2001). In particular, the presence of "somatic mutations" in cancer has been emphasized, where a somatic mutation represents a tumour-specific variant in either the coding or non-coding region that is not present in the normal tissue examined. Furthermore, the detection of a C-tract length polymorphism at position 303 (relative to the Cambridge reference sequence) and the presence of homoplasmy within tumour cells have been promoted as tumorigenic markers.

Nearly all studies published to date attempt to detect the presence of tumour-specific variants by direct sequencing of PCR fragments. This approach, though much more cost-effective and less time consuming than related methods, suffers several shortcomings. The first setback involves the nature of the mtDNA itself. PCR involves the amplification of DNA derived from a heterogeneous population of cells. In the case of PCR amplification of nuclear DNA, both copies of DNA in a cell will either be identical or found at a 1:1 wildtype to mutant ratio. When amplifying several cells, the diploid nature of the human cell ensures that the nuclear DNA sequence of each cell is identical and not likely to fluctuate, barring mutagenic events. The same luxury is not available when assaying mtDNA for variants, as mtDNA is present in tens of thousands of copies per cell, and the relative proportions and populations may fluctuate from generation to generation. Even the study of mtDNA at the single cell level does not resolve the difficulty of determining sequence variation at the level of an individual mitochondrial genome. The second setback involves the sensitivity of PCR-based

mutation detection schemes. Although PCR-based sequencing is suitable for detection of nuclear DNA mutations by virtue of its minimum 1:1 ratio of wildtype to mutant DNA copies, it has recently been reported that the limit of detection for mtDNA variants may be as high as 30%, indicating that mutations or polymorphisms with a frequency of less than 30% in a heterogeneous DNA mixture will not be identified (Tully *et al.*, 1999).

Due to the inherent difficulty in the study of mtDNA sequences and the high limit of detection associated with PCR-based sequencing, a subcloning-based sequencing method, in which 10 individual subclones were sequenced, was employed to not only increase the limit of detection but also to quantify the relative levels of the detected polymorphisms within the samples examined. This approach has never been utilized in a published study of mtDNA variation and cancer. As this approach is very labour-intensive, a short region of mtDNA was chosen for the purpose of this study. The D-loop control region has been reported in several publications as a primary site of mutation or polymorphism in cancer samples, and the C303 tract, which has received special attention, also resides within the D-loop (Bianchi *et al.*, 2001; Penta *et al.*, 2001). More importantly, as the D-loop contains regulatory elements such as the heavy strand origin of replication and both heavy and light strand promoters, sequence variation in this region may lead to downstream effects on replication and transcription. As such, two regions were chosen for study, one within each of HVI and HVII; the HVI PCR primer set was designed to encompass the most frequently polymorphic sites within the mtDNA genome (as determined by MitBase database analysis), and the HVII primer set was designed to encompass the C303 region. Ideally, the study of these regions would serve three purposes: first, detection of polymorphisms in the samples, in order to determine

whether any DNA variants within these regions show a correlation with cancer status; second, determination of the frequency of the polymorphisms detected and the level of homoplasmy; and third, examination of the C303 tract length in order to determine its association with cancer status.

Because the primary stage of the subcloning-based sequencing approach involves PCR amplification, a number of factors were taken into consideration regarding PCR design. The number of rounds of PCR was limited to 30 to decrease the probability of shifting the frequency of mtDNA populations throughout the course of the PCR. Furthermore, the primers' specificity for mtDNA was verified through their inability to amplify DNA from a ρ^0 fibroblast cell line which lacks mtDNA. Only primers capable of amplifying mtDNA specifically were used in further study. When designing the PCR primers to be used for amplification, an effort was made to exclude known polymorphic sites from the primer sequence. As the HV regions are highly polymorphic and the Mitomap database reports certain regions containing one polymorphism at every nucleotide, it was impossible to design 17-20 bp PCR primers that did not overlap with several potentially polymorphic sites. Thus, the PCR primers were designed so as not to overlap with frequently polymorphic sites and in such a way that the 3' tail of each PCR primer does not terminate in proximity to a known polymorphism. However, despite attempts to minimize the influence of polymorphisms on the PCR process, the possibility of a template bias, caused by preferential primer binding and subsequent amplification of certain polymorphic sequences, cannot be excluded in this study.

In order to address whether 10 subclones would provide sufficient insight into the sequence variants present in mitochondrial genomes, a normal blood control sample was

examined by first sequencing 10 subclones. 40 additional subclones were then sequenced in order to determine whether the presence and relative frequency of polymorphisms detected varied considerably with increased subclone sample size (Figure 3-2). Although minor variants were observed when 50 subclones were sequenced, all polymorphisms with a frequency of 2/10 or greater were replicated. The profile of the C303 tract was retained, with minor variations in the percentage of subclones displaying C8, C9 and C10. From these results, I concluded that the results from 10 subclones would sufficiently represent the mtDNA from the sample. As 10 or more subclones were sequenced for each sample, a polymorphism detected in one subclone reflects an approximate frequency of 10% in the analyzed sample, giving a limit of detection of approximately 10% for this study. Although sequencing 10 subclones is sufficient for the purposes of this study, it is not sufficient to detect all polymorphisms or mtDNA populations, as rare variants will likely be overlooked.

Furthermore, to assess the frequency of PCR misincorporation by Taq DNA polymerase, several samples were re-amplified with AmpliTaq Gold, which possesses a low sequence error rate due to the presence of 3' to 5' exonuclease proofreading activity. Upon comparison of the sequence results obtained from amplifications performed with standard Taq and ABI AmpliTaq Gold polymerases, the probability of PCR errors appeared low, and unlikely to significantly affect the results of this study.

There is an ongoing debate in the literature regarding the natural state of the mtDNA in normal individuals. It is generally believed that mtDNA mutation within the coding region is detrimental and the prevalence of sequence variants in individuals unaffected by mtDNA disorders is not widely reported. Furthermore, a debate still exists

regarding the level of heteroplasmy in normal individuals. While historically it has been thought that heteroplasmy was limited to those with mitochondrial disorders, recent data suggest that the level of heteroplasmy, especially in non-coding regions, in normal individuals may in fact be higher than previously reported (Kirches *et al.*, 2001; Monnat *et al.*, 1985). Furthermore, the question of how the overall level of heteroplasmy varies with age remains unclear, as publications suggesting that the level of heteroplasmy remains constant over the lifespan of an individual coexist with reports of heteroplasmy decreasing with age, eventually leading to homoplasmy in very aged individuals (Chomyn and Attardi, 2003; Lagerstrom-Fermer *et al.*, 2001). Before studying mtDNA effects in cancer, it is absolutely essential to determine the levels of the purported tumorigenic markers, including mtDNA polymorphisms, homoplasmy, and C-tract length variation, in normal individuals. No study published to date examines the level of C-tract variation in the normal population, and all published controls involve a comparison tumour DNA to the DNA obtained from adjacent normal tissues.

The results of my study indicate that the HV regions of normal samples are very polymorphic, showing a frequency of up to one polymorphism every 50 bp for HVI and every 100 bp in HVII (Figure 3-1). Although the HVI primers were designed to encompass the most highly polymorphic region of the genome and gave expected results, the high degree of polymorphism in HVII was surprising, as initially this region was intended to assay only C-tract length at position 303. These results indicate that polymorphisms are highly prevalent across both HV regions, albeit to a greater extent in HVI.

As a consequence of the subcloning-based method utilized in this study, it was also possible to determine the associations between certain polymorphisms, allowing for the resolution of clonal populations within samples. A subset of the polymorphic patterns observed in more than one sample likely represents mtDNA haplotypes. Although one would expect that the highly mutagenic nature of the mitochondrion would promote random mutagenesis and thus the formation of several random populations, in reality, the opposite seems to be the norm. Typically, it appears that the mtDNA is limited to two to three populations. Nonetheless, several samples displayed more than three populations, and one sample displayed nine different populations. It must be emphasized that due to the number of subclones sequenced and the limited region of mtDNA examined, these values likely represent the minimum number of populations present within a sample. For instance, only 12 subclones were sequenced for the sample with nine different clonal populations; had more subclones been examined, additional clonal populations may have emerged.

Given that the number of populations within a sample was higher than expected, the corresponding level of heteroplasmy for normal samples was surprisingly high, with 85% of the 19 normal samples showing heteroplasmy (Table 3-2). Although not all samples used in this study were obtained from microdissected tissues, and may therefore contain a heterogeneous mixture of cell and tissue types, the level of homoplasmy or heteroplasmy did not appear to depend on the method with which sample was obtained. Each sample was comprised primarily of the tissue of interest and likely, the sensitivity of the subcloning-based sequencing approach is not sufficient to observe very subtle sequence variation. As such, the levels of heteroplasmy reported here are likely

underestimates, as only approximately 600 bp of sequence was sequenced and nearly 16 kb of the mtDNA is left unexamined. My data therefore support more recent studies examining heteroplasmy within the non-coding region of the mtDNA that report high levels of heteroplasmy in different tissues of normal individuals (Comas *et al.*, 1995; Kirches *et al.*, 2001).

Mitochondrial DNA and Cancer

Given their roles in apoptosis and aerobic metabolism, it seems likely that mitochondria play a central role in determining the balance between cancer prevention and progression. Recent reports have emphasized the potential role of mtDNA in this process by examining the role of "somatic mutations" tumours. The presence of tumour-specific variants, often determined to be homoplasmic, has been reported for several different cancers. As such, a similar approach was adopted to determine the level of polymorphism and heteroplasmy in endometrial carcinomas and multiple myeloma, using the levels obtained from the study of normal individuals as a basis of comparison. At the outset of experimentation, there were no published reports examining the frequency of these markers in either cancer type.

To address the frequency of tumour-specific variation, five pairs of matched endometrial tumour and normal blood samples, derived from the same individual, were sequenced (Table 3-3, Table 3-4). Across the two HV regions, several polymorphisms were detected which were present in either the tumour alone (tumour-specific) or in the normal blood alone (normal-specific). There was no propensity for tumour-specific variants alone; instead, tumour-specific and normal-specific variants occurred with approximately equal frequency. Most frequently, the unique variant occurred in a single

subclone, but one polymorphism appeared in 7/11 tumour subclones in a sample. There was no apparent trend to distinguish tumour samples from their normal counterparts. It should be noted that different tissues and organs, even within a single individual, can show sequence variation (Kirches *et al.*, 2001). Thus, tumour-specific variants must be rigorously studied to ensure that what appears to be a tumour-specific polymorphism or mutation is not simply the result of tissue- or organ-specific differentiation.

A total of 76 and 55 polymorphisms were detected in endometrial and multiple myeloma samples, respectively (Figure 3-1). Although the multiple myeloma samples exhibited an elevated mutation frequency, due to the small sample size of the data set the relative increase is statistically insignificant when compared to the frequency of polymorphism in normal samples. No single polymorphism or group of polymorphisms appears to distinguish either cancer type from the normal controls (Figure 3-1). In addition, there was no statistical correlation between the frequency of polymorphism and cancer grade, indicating that individuals with more advanced cancers do not necessarily display higher or lower levels of polymorphism than those with less advanced cancers.

It has been suggested that polymorphism at nt 146, 150, 152, 189, 195, and 408 may have functional significance, as they directly lie within origin and promoter sequences (Chomyn and Attardi, 2003). The polymorphism at site 408 is outside the region studied here, and therefore its frequency was not determined. While no polymorphisms were detected at nt 146, 150 and 152, polymorphisms were detected at nt 189 and 195. A189C was detected in a single endometrial cancer sample, AH221, a sample lacking MSI and displaying the highest number of sequence polymorphisms of all endometrial samples studied. T195C, also within the regulatory region, was detected in

two endometrial samples, one multiple myeloma sample, and five normal controls. Interestingly, this polymorphism was found within one of the matched pairs of samples, and was found within the normal blood sample, but not within the tumour, indicating a normal-specific variant. Due to its prevalence in normal samples, it is unlikely that the T195C polymorphism has a functional significance in the samples studied.

The T16189C polymorphism is reported by Liu *et al.* to be associated with susceptibility to endometrial cancer (Liu *et al.*, 2003a). In this study, this polymorphism was found in 13 different samples, two of which were endometrial samples, four being multiple myeloma samples, and seven of which were normal controls. Interestingly, of the seven normal samples displaying polymorphism at position 16189, all but two were homoplasmic for the polymorphism, whereas none of the cancer samples displaying the polymorphism were homoplasmic at this site. It is therefore very unlikely that this site serves as a susceptibility marker for endometrial carcinoma.

No correlation between patient age and number of D-loop polymorphisms was observed for endometrial cancer patients. Interestingly, however, the multiple myeloma samples exhibited a negative correlation between patient age and the frequency of D-loop polymorphism. As such, younger patients display a statistically larger number of polymorphisms than older patients in this region of the mitochondrial genome. Reports in the literature suggest the opposite association, with more aged individuals exhibiting an increased number of polymorphisms (Michikawa *et al.*, 1999). The increased incidence of mutation is attributed to the observation that levels of oxidative damage tend to increase with age, likely due to accumulation of reactive oxygen species, such as hydrogen peroxide and hydroxyl radicals formed during oxidative respiration (Chomyn

and Attardi, 2003). It is possible that patients affected by multiple myeloma at a younger age possess higher levels of reactive oxygen species than their normal counterparts, thereby allowing the accumulation of mutations and polymorphisms over a shorter period of time. This observation, however, cannot account for the low frequency of polymorphism in more aged multiple myeloma patients. Recent literature also reports an increase in the level of cellular homoplasmy over time (Chomyn and Attardi, 2003). While this observation may not entirely account for the decrease in mutation frequency in older patients, it is possible that older multiple myeloma patients tend to become more homoplasmic with time. Interestingly, the sole homoplasmic multiple myeloma patient was also the oldest. Further investigation into the frequency of polymorphism and the level of homoplasmy in a greater number of samples is required before any conclusions can be made, as the sample size of the multiple myeloma patient set is limited and the observations may be attributable to sampling error.

There has been little investigation of the relationship of nuclear microsatellite instability – indicating deficient nuclear mismatch repair – and the presence of mtDNA mutations. One study examined the presence of mtDNA mutations within the D-loop and five mitochondrial protein coding genes in gastric carcinoma, and determined that there was no significant relationship between nuclear MSI status and the frequency of mtDNA mutations. There was, however, an inverse correlation between the presence of large mtDNA deletions and tumours lacking nuclear MSI (Maximo *et al.*, 2001). As no large deletions were detected in this study, their correlation to MSI status could not be determined here. However, for endometrial cancer, it appears that samples exhibiting nuclear MSI possessed less than half the average number of polymorphisms of samples

lacking nuclear MSI (4.6 vs. 10.4) ($p = 0.0122$). It is unclear what may cause this inverse relationship, although it may indicate an association between the mitochondria and the nucleus that has not yet been explored.

Several unreported polymorphisms were detected in this study (Table 3-1). Reamplification of certain samples with AmpliTaq Gold allowed for confirmation of the presence of sequence variants. In most cases, the sequence variation detected with normal Taq polymerase mirrored the results obtained using AmpliTaq Gold suggesting that the frequency of Taq polymerase error in this study is likely very low. Several of the undetected polymorphisms appear in more than one sample or more than one subclone, appearing at times at homoplasmic levels. 13 of the 21 unreported polymorphisms detected are transitions, consistent with the deamination of cytosine residues. Reactive oxygen species are capable of deaminating cytosine, leaving a uracil in its place. During subsequent rounds of mtDNA replication, the uracil will mispair with an adenine, leaving a G to A transition and associated C to T transition. Most of the remaining unreported polymorphisms are C to A transversions. These errors stem from oxidative damage leading to the formation of 8-hydroxyguanine, which subsequently mispairs with adenine during DNA replication leading to C to A transversions (Cheng *et al.*, 1992). This further suggests that the unreported polymorphisms detected here are legitimate and may represent minor populations within the samples.

Recent data suggest that chemotherapeutic agents may induce mtDNA damage, indicated by a higher number of heteroplasmic mtDNA mutations in patients undergoing chemotherapy for chronic lymphocytic leukemia (Carew *et al.*, 2003). Statistical analysis could not be performed to determine the effect of cancer treatment on the endometrial

cancer and multiple myeloma samples employed in this study as the number of treated patients was too low. Additional sequence data obtained from treated individuals would be necessary to determine the effect of chemotherapy on frequency of mutation in these cancers. In order to study the effect of chemotherapy, a longitudinal study involving several patients would be necessary. The mtDNA variation would be determined at the time of diagnosis, and monitored throughout the progression of their cancer and resulting treatment. The degree of sequence variation would ideally be observed in single individuals, as it is impossible to control for the natural variation that occurs between individuals.

The frequency of homoplasmy in cancer samples is approximately equal to that observed in normal controls (Table 3-2). Although endometrial samples exhibited a level of homoplasmy slightly higher than the normal samples, and the multiple myeloma samples exhibited slightly lower levels of homoplasmy, the level of homoplasmy is statistically indistinguishable between the three classes of samples examined. All homoplasmic samples contained at least one sequence variant compared to the Cambridge reference sequence, regardless of whether the sample was derived from tumour or normal DNA. It therefore appears that the presence of homoplasmy within a sample is not necessarily indicative of cancer status, and as such, would be ineffective as a marker of tumourogenesis or carcinogenesis.

C303 Expansion as a Carcinogenic Marker

Mitochondrial DNA could serve as an ideal candidate for a carcinogenic marker due to its high copy number and ease of detection in bodily fluids, thus alleviating invasive biopsy procedures (Fliss *et al.*, 2000). One such marker has been recently

proposed: the expansion of a C-tract at position 303, apparently inherent to tumour cells. The prevalence of this marker in the general population had not previously been published. Although the Cambridge reference sequence indicates that a C-tract length of seven cytosines at position 303 is typical in normal individuals, upon examination of this locus in 19 different normal individuals, 54% of the samples displayed an expansion of the C303 tract above the normal C7 length (Figure 3-3). C303 tract lengths varied between seven and 10 cytosines and no subclones exhibited a length outside this reported range. Importantly, up to four different C303 tract expansions were often detected within individual normal controls, indicating that expansion at this locus likely occurs frequently within normal individuals.

The C303 tract length in cancer samples varied from C7 to C10, similar to the expansion observed in normal individuals (Figure 3-3). Upon statistical analysis, it was determined that normal samples expand at a frequency approximately equal to the level of expansion observed in cancer samples. The only statistical deviation was observed in regard to the C9 length, in which cancer samples had a statistically higher proportion of C9 lengths than normal samples. The increase in C9 subclones can likely be attributed to the increased level of C9 variants in multiple myeloma subclones. A functional relevance of the increase to C9 is unknown as the role of the C-tract in mtDNA replication and/or transcription remains unclear despite much speculation regarding its function. The C-tract is located within a conserved sequence block which, together with two other conserved sequence blocks, is believed to play a role in the formation of the R-loop and subsequent initiation of replication from the heavy strand. It has been suggested that significant alteration of the C-tract could hinder initiation of transcription, thereby

providing a means of negative selection of such DNA molecules (Sanchez-Cespedes *et al.*, 2001). Due to the prevalence of C-tract expansion in normal individuals, and given that the mutation spectrum in normal individuals is nearly identical to reported tumour-associated expansions in terms of possible C-tract lengths, it appears unlikely that expansion at this site would significantly impair mitochondrial function or play a role in pathology.

It seems unlikely that the C303 length variation observed in normal individuals is the result of PCR or sequencing slippage, as C-tract length results are reproducible when samples are reamplified by PCR and subjected to further sequence analysis. Furthermore, a tract of 6 cytosines follows the C303 tract, with the two tracts separated by a single thymine. If the expansions observed were due to sequencing or PCR errors, one would expect that the tract of six cytosines would be affected as well. However, all 1025 subclones sequenced exhibited a tract of six cytosines at this position. The reproducibility of sequence results, combined with the lack of variation at this adjacent cytosine tract and the lack of alteration at the single thymine that separates the two tracts, strongly suggest that the C303 variation observed is not due to PCR or sequencing error.

Microfluidic Chip Assays

Although the means required for more precise study of mtDNA is not yet available, as technological methods become more advanced, our understanding of the role of mitochondria and mtDNA in carcinogenesis and tumourogenesis will become more refined. The analysis of mtDNA using microfluidic chip applications holds the promise of examining mtDNA on the level of a single cell or perhaps even a single mitochondrion. As such, the essential groundwork for such innovative research

techniques has been laid through the work presented in this thesis. Integrated on-chip labeling, restriction enzyme digestion and heteroduplex analysis allows for the rapid detection of heteroplasmy or polymorphisms and represents a significant advance in the microfluidics field. Furthermore, the unique profile of each polymorphic spectrum analyzed to date exhibits the great potential for the use of microfluidic technology for mtDNA mutation detection in the future. The use of DNA labeling dyes that bind single-stranded DNA will allow the further integration of SSCP analysis, pushing the capabilities of such a system even further than is currently feasible by standard means of mutation detection.

Summary

The use of a subcloning-based sequencing method has proven to be a precise approach to delineating mtDNA sequence changes in normal and cancer samples. The increased sensitivity of this approach has allowed the detection of low-frequency polymorphisms and resulting heteroplasmy that may have been undetectable by other experimental means. Furthermore, the prevalence of purported tumorigenic markers in normal individuals has profound implications for the use of these mtDNA markers in prognostic/diagnostic testing. A reliable marker for carcinogenesis should not be present in the normal population. Moreover, it is essential that mitochondrial genetics regarding tissue or organ-specific mtDNA differentiation be taken into consideration when searching for tumour-specific mtDNA mutations. Although there is likely an association between the mitochondria and the prevention or initiation of cancer, through

mitochondrial ties to energy metabolism, apoptosis, and its effect on nuclear gene regulation, it is doubtful that a single mtDNA marker will emerge within the mtDNA control region for the cancers examined here.

In order to obtain a more complete picture regarding the state of the mitochondrial genome and certain aspects of cancer, a more sensitive and cost-effective means of analysis must be employed. As such means become available, an increased number of regions and samples could be analyzed, including a greater number of patients of different ages, undergoing a variety of treatments in order to determine what affect age and treatment may have on the mtDNA. Further investigation of the apparent association between nuclear MSI and mtDNA polymorphism is required in order to validate my results and to explore possible mechanisms for this association. Finally, the development of microfluidic chip technology will allow integration of mutation analysis systems into a less time-consuming, more efficient means of analyzing minute amounts of mtDNA. Studies of mtDNA polymorphism and the prevalence of heteroplasmy in both normal and cancerous scenarios have been hindered by the available molecular approaches, and it is with the development of technologies capable of examining mtDNA at the single genome level that the study of mtDNA in cancer will be definitive.

References

(2003). National Cancer Institute of Canada: Canadian Cancer Statistics 2003 (Toronto, Canada).

Amuthan G, Biswas G, Zhang SY, Klein-Szanto A, Vijayasarathy C, Avadhani NG (2001) Mitochondria-to-nucleus stress signaling induces phenotypic changes, tumor progression and cell invasion. EMBO J *20*:1910-20.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature *290*:457-65.

Anker P, Stroun M (2001) Tumor-related alterations in circulating DNA, potential for diagnosis, prognosis and detection of minimal residual disease. Leukemia *15*:289-91.

Anson RM, Croteau DL, Stierum RH, Filburn C, Parsell R, Bohr VA (1998) Homogenous repair of singlet oxygen-induced DNA damage in differentially transcribed regions and strands of human mitochondrial DNA. Nucleic Acids Res *26*:662-8.

Aquadro C, Greenberg B (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics *103*:287-312.

Attardi G, Ojala D (1971) Mitochondrial ribosome in HeLa cells. Nat New Biol *229*:133-6.

Attardi G, Schatz G (1988) Biogenesis of mitochondria. Annu Rev Cell Biol *4*:289-333.

Backhouse CJ, Gajdal A, Pilarski LM, Crabtree HJ (2003) Improved resolution with microchip-based enhanced field inversion electrophoresis. Electrophoresis *24*:1777-86.

Bandy B, Davison AJ (1990) Mitochondrial mutations may increase oxidative stress: implications for carcinogenesis and aging? Free Radic Biol Med *8*:523-39.

Bergsagel PL, Chesi M, Nardini E, Brents LA, Kirby SL, Kuehl WM (1996) Promiscuous translocations into immunoglobulin heavy chain switch regions in multiple myeloma. Proc Natl Acad Sci U S A *93*:13931-6.

Bianchi NO, Bianchi MS, Richard SM (2001) Mitochondrial genome instability in human cancers. Mutat Res *488*:9-23.

Bogenhagen D, Clayton DA (1974) The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. J Biol Chem *249*:7991-5.

Bogenhagen D, Clayton DA (1977) Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. Cell *11*:719-27.

Bohr VA, Stevnsner T, de Souza-Pinto NC (2002) Mitochondrial DNA repair of oxidative damage in mammalian cells. Gene *286*:127-34.

Brega A, Vesco C (1971) Ribonucleoprotein particles involved in HeLa mitochondrial protein synthesis. Nat New Biol *229*:136-9.

Brown TA, Clayton DA (2002) Release of replication termination controls mitochondrial DNA copy number after depletion with 2',3'-dideoxycytidine. Nucleic Acids Res *30*:2004-10.

Bruin GJ (2000) Recent developments in electrokinetically driven analysis on microfabricated devices. Electrophoresis *21*:3931-51.

Calloway CD, Reynolds RL, Herrin GL, Jr., Anderson WW (2000) The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am J Hum Genet *66*:1384-97.

Carew JS, Huang P (2002) Mitochondrial defects in cancer. Mol Cancer *1*:9.

Carew JS, Zhou Y, Albitar M, Carew JD, Keating MJ, Huang P (2003) Mitochondrial DNA mutations in primary leukemia cells after chemotherapy: clinical significance and therapeutic implications. Leukemia *17*:1437-47.

Chang DD, Clayton DA (1984) Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell *36*:635-43.

Chang DD, Clayton DA (1987) A mammalian mitochondrial RNA processing activity contains nucleus-encoded RNA. Science *235*:1178-84.

Chen X, Prosser R, Simonetti S, Sadlock J, Jagiello G, Schon EA (1995) Rearranged mitochondrial genomes are present in human oocytes. Am J Hum Genet *57*:239-47.

Cheng K, Cahill D, Kasai H, Nishimura S, Loeb L (1992) 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G to T and A to C substitutions. J Biol Chem. *267*:166-172.

Chinnery PF (2002) Modulating heteroplasmy. Trends Genet *18*:173-6.

Chinnery PF, Thorburn DR, Samuels DC, White SL, Dahl HM, Turnbull DM, Lightowlers RN, Howell N (2000) The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? Trends Genet *16*:500-5.

Chomyn A, Attardi G (2003) MtDNA mutations in aging and apoptosis. Biochem Biophys Res Commun *304*:519-29.

Clayton DA (1982) Replication of animal mitochondrial DNA. Cell *28*:693-705.

Clayton DA (1984) Transcription of the mammalian mitochondrial genome. Annu Rev Biochem *53*:573-94.

Clayton DA (1992) Transcription and replication of animal mitochondrial DNAs. Int Rev Cytol *141*:217-32.

Coller HA, Khrapko K, Bodyak ND, Nekhaeva E, Herrero-Jimenez P, Thilly WG (2001) High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. Nat Genet *28*:147-50.

Comas D, Paabo S, Bertranpetit J (1995) Heteroplasmy in the control region of human mitochondrial DNA. Genome Res *5*:89-90.

Croteau DL, ap Rhys CM, Hudson EK, Dianov GL, Hansford RG, Bohr VA (1997) An oxidative damage-specific endonuclease from rat liver mitochondria. J Biol Chem *272*:27338-44.

Croteau DL, Stierum RH, Bohr VA (1999) Mitochondrial DNA repair pathways. Mutat Res *434*:137-48.

DiMauro S, Schon EA (2001) Mitochondrial DNA mutations in human disease. Am J Med Genet *106*:18-26.

Driggers WJ, Holmquist GP, LeDoux SP, Wilson GL (1997) Mapping frequencies of endogenous oxidative damage and the kinetic response to oxidative stress in a region of rat mtDNA. Nucleic Acids Res *25*:4362-9.

Dujon B, Slonimski PP, Weill L (1974) Mitochondrial genetics IX: A model for recombination and segregation of mitochondrial genomes in saccharomyces cerevisiae. Genetics *78*:415-37.

Fan L, Sanschagrin PC, Kaguni LS, Kuhn LA (1999) The accessory subunit of mtDNA polymerase shares structural homology with aminoacyl-tRNA synthetases: implications for a dual role as a primer recognition factor and processivity clamp. Proc Natl Acad Sci U S A *96*:9527-32.

Fernandez-Silva P, Enriquez JA, Montoya J (2003) Replication and transcription of mammalian mitochondrial DNA. Exp Physiol *88*:41-56.

Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM, Jen J, Sidransky D (2000) Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. Science *287*:2017-9.

Footz T, Wunsam S, Kulak S, Crabtree HJ, Glerum DM, Backhouse CJ (2001) Sample purification on a microfluidic device. Electrophoresis *22*:3868-75.

Ghivizzani SC, Madsen CS, Nelen MR, Ammini CV, Hauswirth WW (1994) *In organello* footprint analysis of human mitochondrial DNA: human mitochondrial transcription factor A interactions at the origin of replication. Mol Cell Biol *14*:7717-30.

Graves SW, Johnson AA, Johnson KA (1998) Expression, purification, and initial kinetic characterization of the large subunit of the human mitochondrial DNA polymerase. Biochemistry *37*:6050-8.

Gray H, Wong TW (1992) Purification and identification of subunit structure of the human mitochondrial DNA polymerase. J Biol Chem *267*:5835-41.

Ha PK, Tong BC, Westra WH, Sanchez-Cespedes M, Parrella P, Zahurak M, Sidransky D, Califano JA (2002) Mitochondrial C-tract alteration in premalignant lesions of the head and neck: a marker for progression and clonal proliferation. Clin Cancer Res *8*:2260-5.

Hammarsund M, Wilson W, Corcoran M, Merup M, Einhorn S, Grander D, Sangfelt O (2001) Identification and characterization of two novel human mitochondrial elongation factor genes, *hEFG2* and *hEFG1*, phylogenetically conserved through evolution. Hum Genet *109*:542-50.

Hixson JE, Wong TW, Clayton DA (1986) Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J Biol Chem *261*:2384-90.

Igney FH, Krammer PH (2002) Death and anti-death: tumour resistance to apoptosis. Nat Rev Cancer *2*:277-88.

Jazin EE, Cavelier L, Eriksson I, Oreland L, Gyllensten U (1996) Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. PNAS *93*:12382-12387.

Jenuth JP, Peterson AC, Fu K, Shoubridge EA (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet *14*:146-51.

Johns DR (1995) Seminars in medicine of the Beth Israel Hospital, Boston. Mitochondrial DNA and disease. N Engl J Med *333*:638-44.

Johnson AA, Tsai Y, Graves SW, Johnson KA (2000) Human mitochondrial DNA polymerase holoenzyme: reconstitution and characterization. Biochemistry *39*:1702-8.

Kang D, Nishida J, Iyama A, Nakabeppu Y, Furuichi M, Fujiwara T, Sekiguchi M, Takeshige K (1995) Intracellular localization of 8-oxo-dGTPase in human cells, with special reference to the role of the enzyme in mitochondria. J Biol Chem *270*:14659-65.

Khrapko K, Coller HA, Andre PC, Li XC, Hanekamp JS, Thilly WG (1997) Mitochondrial mutational spectra in human cells and tissues. Proc Natl Acad Sci USA *94*:13798-803.

Kirches E, Michael M, Warich-Kirches M, Schneider T, Weis S, Krause G, Mawrin C, Dietzmann K (2001) Heterogeneous tissue distribution of a mitochondrial DNA polymorphism in heteroplasmic subjects without mitochondrial disorders. J Med Genet *38*:312-7.

Kuehl WM, Bergsagel PL (2002) Multiple myeloma: evolving genetic events and host interactions. Nat Rev Cancer *2*:175-87.

Lagerstrom-Fermer M, Olsson C, Forsgren L, Syvanen AC (2001) Heteroplasmy of the human mtDNA control region remains constant during life. Am J Hum Genet *68*:1299-301.

Lakshmipathy U, Campbell C (1999) Double strand break rejoining by mammalian mitochondrial extracts. Nucleic Acids Res *27*:1198-204.

Larsson NG, Wang J, Wilhelmsson H, Oldfors A, Rustin P, Lewandoski M, Barsh GS, Clayton DA (1998) Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. Nat Genet *18*:231-6.

Lecrenier N, Foury F (2000) New features of mitochondrial DNA replication system in yeast and man. Gene *246*:37-48.

Lee DY, Clayton DA (1996) Properties of a primer RNA-DNA hybrid at the mouse mitochondrial DNA leading-strand origin of replication. J Biol Chem *271*:24262-9.

Lee DY, Clayton DA (1998) Initiation of mitochondrial DNA replication by transcription and R-loop processing. J Biol Chem *273*:30614-21.

Lee HC, Pang CY, Hsu HS, Wei YH (1994) Ageing-associated tandem duplications in the D-loop of mitochondrial DNA of human muscle. FEBS Lett *354*:79-83.

Leonard JV, Schapira AH (2000) Mitochondrial respiratory chain disorders I: mitochondrial DNA defects. Lancet *355*:299-304.

Liao HX, Spremulli LL (1989) Interaction of bovine mitochondrial ribosomes with messenger RNA. J Biol Chem *264*:7518-22.

Liao HX, Spremulli LL (1991) Initiation of protein synthesis in animal mitochondria. Purification and characterization of translational initiation factor 2. J Biol Chem *266*:20714-9.

Lightowlers RN, Chinnery PF, Turnbull DM, Howell N (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. Trends Genet *13*:450-5.

Liu VW, Wang Y, Yang HJ, Tsang PC, Ng TY, Wong LC, Nagley P, Ngan HY (2003a) Mitochondrial DNA variant 16189T>C is associated with susceptibility to endometrial cancer. Hum Mutat *22*:173-4.

Liu VW, Yang HJ, Wang Y, Tsang PC, Cheung AN, Chiu PM, Ng TY, Wong LC, Nagley P, Ngan HY (2003b) High frequency of mitochondrial genome instability in human endometrial carcinomas. Br J Cancer *89*:697-701.

Lu H, Forbes RA, Verma A (2002) Hypoxia-inducible factor 1 activation by aerobic glycolysis implicates the Warburg effect in carcinogenesis. J Biol Chem *277*:23111-5.

Manfredi G, Thyagarajan D, Papadopoulou LC, Pallotti F, Schon EA (1997) The fate of human sperm-derived mtDNA in somatic cells. Am J Hum Genet *61*:953-60.

Martens PA, Clayton DA (1979) Mechanism of mitochondrial DNA replication in mouse L-cells: localization and sequence of the light-strand origin of replication. J Mol Biol *135*:327-51.

Mason PA, Matheson EC, Hall AG, Lightowlers RN (2003) Mismatch repair activity in mammalian mitochondria. Nucl. Acids. Res. *31*:1052-1058.

Maximo V, Soares P, Seruca R, Rocha AS, Castro P, Sobrinho-Simoes M (2001) Microsatellite instability, mitochondrial DNA large deletions, and mitochondrial DNA mutations in gastric carcinoma. Genes Chromosomes Cancer *32*:136-43.

Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics *152*:1103-10.

Michikawa Y, Mazzucchelli F, Bresolin N, Scarlato G, Attardi G (1999) Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. Science *286*:774-9.

Moore DS, McCabe GP (1993). Introduction to the practice of statistics. 2nd ed. New York: W.H. Freeman and Company.

Monnat RJ, Jr., Maxwell CL, Loeb LA (1985) Nucleotide sequence preservation of human leukemic mitochondrial DNA. Cancer Res *45*:1809-14.

Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G (1982) Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc Natl Acad Sci USA *79*:7195-9.

Montoya J, Gaines GL, Attardi G (1983) The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell *34*:151-9.

Montoya J, Ojala D, Attardi G (1981) Distinctive features of the 5'-terminal sequences of the human mitochondrial mRNAs. Nature *290*:465-70.

Moraes CT, Kenyon L, Hao H (1999) Mechanisms of human mitochondrial DNA maintenance: The determining role of primary sequence and length over function. Mol. Biol. Cell *10*:3345-3356.

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Mol Biol Evol *18*:1833-7.

Murdock DG, Christacos NC, Wallace DC (2000) The age-related accumulation of a mitochondrial DNA control region mutation in muscle, but not brain, detected by a sensitive PNA-directed PCR clamping based method. Nucleic Acids Res *28*:4350-5.

Murphy WI, Attardi B, Tu C, Attardi G (1975) Evidence for complete symmetrical transcription in vivo of mitochondrial DNA in HeLa cells. J Mol Biol *99*:809-14.

Myers KA, Saffhill R, O'Connor PJ (1988) Repair of alkylated purines in the hepatic DNA of mitochondria and nuclei in the rat. Carcinogenesis *9*:285-92.

Nass MM (1969) Mitochondrial DNA. I. Intramitochondrial distribution and structural relations of single- and double-length circular DNA. J Mol Biol *42*:521-8.

Neupert W (1997) Protein import into mitochondria. Annu Rev Biochem *66*:863-917.

Nilsen H, Otterlei M, Haug T, Solum K, Nagelhus TA, Skorpen F, Krokan HE (1997) Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. Nucleic Acids Res *25*:750-5.

Obeid PJ, Christopoulos TK, Crabtree HJ, Backhouse CJ (2003) Microfabricated device for DNA and RNA amplification by continuous-flow polymerase chain reaction and reverse transcription-polymerase chain reaction with cycle number selection. Anal Chem *75*:288-95.

Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature *290*:470-4.

Olivo PD, Van de Walle MJ, Laipis PJ, Hauswirth WW (1983) Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature *306*:400-2.

Ono T, Isobe K, Nakada K, Hayashi JI (2001) Human cells are protected from mitochondrial dysfunction by complementation of DNA products in fused mitochondria. Nat Genet *28*:272-5.

Otterlei M, Haug T, Nagelhus TA, Slupphaug G, Lindmo T, Krokan HE (1998) Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. Nucleic Acids Res *26*:4611-7.

Parfait B, Rustin P, Munnich A, Rotig A (1998) Co-amplification of nuclear pseudogenes and assessment of heteroplasmy of mitochondrial DNA mutations. Biochem Biophys Res Comm *247*:57-9.

Parrella P, Seripa D, Matera MG, Rabitti C, Rinaldi M, Mazzarelli P, Gravina C, Gallucci M, Altomare V, Flammia G, Casalino B, Benedetti-Panici PL, Fazio VM (2003) Mutations of the D310 mitochondrial mononucleotide repeat in primary tumors and cytological specimens. Cancer Lett *190*:73-7.

Parrella P, Xiao Y, Fliss M, Sanchez-Cespedes M, Mazzarelli P, Rinaldi M, Nicol T, Gabrielson E, Cuomo C, Cohen D, Pandit S, Spencer M, Rabitti C, Fazio VM, Sidransky D (2001) Detection of mitochondrial DNA mutations in primary breast cancer and fine-needle aspirates. Cancer Res *61*:7623-6.

Penta JS, Johnson FM, Wachsman JT, Copeland WC (2001) Mitochondrial DNA in human malignancy. Mutat Res *488*:119-33.

Pinz KG, Shibutani S, Bogenhagen DF (1995) Action of mitochondrial DNA polymerase gamma at sites of base loss or oxidative damage. J Biol Chem *270*:9202-6.

Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD, Trush MA, Kinzler KW, Vogelstein B (1998) Somatic mutations of the mitochondrial genome in human colorectal tumours. Nat Genet *20*:291-3.

Robertson G (2003) Screening for endometrial cancer. Med J Aust *178*:657-9.

Sanchez-Cespedes M, Parrella P, Nomoto S, Cohen D, Xiao Y, Esteller M, Jeronimo C, Jordan RC, Nicol T, Koch WM, Schoenberg M, Mazzarelli P, Fazio VM, Sidransky D (2001) Identification of a mononucleotide repeat as a major target for mitochondrial DNA alterations in human tumors. Cancer Res *61*:7015-9.

Satoh M, Kuroiwa T (1991) Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp Cell Res *196*:137-40.

Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. N Engl J Med *347*:576-80.

Schwickert G, Walenta S, Sundfor K, Rofstad EK, Mueller-Klieser W (1995) Correlation of high lactate levels in human cervical cancer with incidence of metastasis. Cancer Res *55*:4757-9.

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem *66*:409-35

Sherman ME (2000) Theories of Endometrial Carcinogenesis: A Multidisciplinary Approach. Mod Pathol *13*:295-308

Shigenaga MK, Hagen TM, Ames BN (1994) Oxidative damage and mitochondrial decay in aging. Proc Natl Acad Sci U S A *91*:10771-8.

Shitara H, Kaneda H, Sato A, Inoue K, Ogura A, Yonekawa H, Hayashi J-I (2000) Selective and continuous elimination of mitochondria microinjected into mouse eggs from spermatids, but not from liver cells, occurs throughout embryogenesis. Genetics *156*:1277-1284.

Shoubridge EA, Karpati G, Hastings KE (1990) Deletion mutants are functionally dominant over wild-type mitochondrial genomes in skeletal muscle fiber segments in mitochondrial disease. Cell *62*:43-9.

Simon DK, Johns DR (1999) Mitochondrial disorders: clinical and genetic features. Annu Rev Med *50*:111-27.

Smeitink J, van den Heuvel L, DiMauro S (2001) The genetics and pathology of oxidative phosphorylation. Nat Rev Genet *2*:342-52.

St John J, Sakkas D, Dimitriadi K, Barnes A, Maclin V, Ramey J, Barratt C, De Jonge C (2000) Failure of elimination of paternal mitochondrial DNA in abnormal embryos. Lancet *355*:200.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (1999) Ubiquitin tag for sperm mitochondria. Nature *402*:371-2.

t Hart LM, Jansen JJ, Lemkes HH, de Knijff P, Maassen JA (1996) Heteroplasmy levels of a mitochondrial gene mutation associated with diabetes mellitus decrease in leucocyte DNA upon aging. Hum Mutat *7*:193-7.

Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta *1410*:103-23.

Takao M, Aburatani H, Kobayashi K, Yasui A (1998) Mitochondrial targeting of human DNA glycosylases for repair of oxidative DNA damage. Nucleic Acids Res *26*:2917-22.

Tang Y, Manfredi G, Hirano M, Schon EA (2000) Maintenance of human rearranged mitochondrial DNAs in long-term cultured transmitochondrial cell lines. Mol Biol Cell *11*:2349-58.

Tapper DP, Clayton DA (1981) Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands. J Biol Chem *256*:5109-15.

Tiranti V, Savoia A, Forti F, D'Apolito MF, Centra M, Rocchi M, Zeviani M (1997) Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. Hum Mol Genet *6*:615-25.

Tully LA, Schwarz FP, Levin BC (1999) Development of a heteroplasmic mitochondrial DNA standard reference material for detection of heteroplasmy and low frequency mutations Proceedings of the 10th International Symposium in Human Identification, Orlando, FL URL: http://www.promega.com/geneticidproc/ussymp10proc/abstracts/24Tully.pdf

Wakeley J (1994) Substitution-rate variation among sites and the estimation of transition bias. Mol Biol Evol *11*:436-442.

Walberg MW, Clayton DA (1983) In vitro transcription of human mitochondrial DNA. Identification of specific light strand transcripts from the displacement loop region. J Biol Chem *258*:1268-75.

Walenta S, Salameh A, Lyng H, Evensen JF, Mitze M, Rofstad EK, Mueller-Klieser W (1997) Correlation of high lactate levels in head and neck tumors with incidence of metastasis. Am J Pathol *150*:409-15.

Wallace DC (1982) Structure and evolution of organelle genomes. Microbiol Rev *46*:208-40.

Wallace DC (1992a) Diseases of the mitochondrial DNA. Annu Rev Biochem *61*:1175-212

Wallace DC (1992b) Mitochondrial genetics: a paradigm for aging and degenerative diseases? Science *256*:628-32.

Wallace DC (1999) Mitochondrial diseases in man and mouse. Science *283*:1482-8.

Wallace DC, Lott MT (2003) MITOMAP: A Human Mitochondrial Genome Database. http://www.mitomap.org

Wang J, Silva JP, Gustafsson CM, Rustin P, Larsson NG (2001) Increased in vivo apoptosis in cells lacking mitochondrial DNA gene expression. Proc Natl Acad Sci USA *98*:4038-43.

Warburg O (1956) On the origin of cancer cells. Science *123*:309-14

Wardell TM, Ferguson E, Chinnery PF, Borthwick GM, Taylor RW, Jackson G, Craft A, Lightowlers RN, Howell N, Turnbull DM (2003) Changes in the human mitochondrial genome after treatment of malignant disease. Mutat Res *525*:19-27.

Woischnik M, Moraes CT (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res *12*:885-93.

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. Int Rev Cytol *141*:173-216.

Wong TW, Clayton DA (1985a) In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell *42*:951-8.

Wong TW, Clayton DA (1985b) Isolation and characterization of a DNA primase from human mitochondria. J Biol Chem *260*:11530-5.

Xu B, Clayton DA (1996) RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. EMBO J *15*:3135-43.

Yoneda M, Chomyn A, Martinuzzi A, Hurko O, Attardi G (1992) Marked replicative advantage of human mtDNA carrying a point mutation that causes the MELAS encephalomyopathy. Proc Natl Acad Sci USA *89*:11164-11168.